

Resumen ejecutivo del proyecto
Escuela de Organización Industrial
Programa Superior en Big Data & Analytics
Madrid. Mayo 2016



FITNESSINTELLIGENCE

CONTAMOS CON LOS MEJORES PROFESIONALES PARA LLEVAR
ADELANTE NUESTRA VISIÓN
NUESTRO EQUIPO



Contacto:

www.fitnessintelligence.eu
contacto@fitnessintelligence.eu

[CONTENIDOS]

3

Misión, visión y resumen conceptual

5

Problema de negocio

12

Fitness Intelligence

20

Producto

23

Datos estructurados

25

Datos no estructurados

29

Arquitectura

33

Business Case

36

Apéndice A

Cambiar la manera en que se concibe y practica el deporte en todo el mundo ayudando a nuestros usuarios a alcanzar el máximo rendimiento físico que deseen y el estado óptimo de salud.



Nuestros clientes: Los centros de fitness. En primera fase, aquellos pertenecientes al segmento low cost.

Nuestros usuarios: Los clientes de nuestros clientes; la gente que acude a entrenar a los gimnasios

Nuestra solución: **Fitness Intelligence**

Lo que queremos es llevar el Internet of Things a todos los centros de fitness. A través de la sensorización intensiva de las máquinas de ejercicio y del seguimiento biométrico de los usuarios, extraemos toda la información sobre lo que está aconteciendo en el gimnasio: Qué ejercicio está realizando cada usuario, qué rendimiento está teniendo durante el ejercicio, cuánto tiempo permanece parado, etc. Con ello, y gracias a algoritmos de machine learning, diseñaremos y modificaremos en tiempo real su rutina de ejercicios para maximizar el cumplimiento de los objetivos de los usuarios, evitar lesiones y optimizar colas dentro del gimnasio, aumentando así la felicidad y la comodidad de los deportistas, que además podrán tener a su disposición en todo momento la información sobre su rutina de ejercicio y sobre su performance a través de su dispositivo móvil y de consolas fijas instaladas en el gimnasio.

Esto servirá para aumentar los ingresos del gimnasio gracias al aumento de la retención y de la adquisición de nuevos clientes captados por los propios usuarios, pero además permitirá conseguir ingresos adicionales a través de la venta de productos alimenticios y bebidas isotónicas o de sesiones de fisioterapia gracias al márketing contextual, que podrán serles recomendados por el sistema de manera automática en los momentos de la verdad (por ejemplo, al finalizar una rutina de alta intensidad).

Por otra parte, dotaremos al gimnasio cliente de Fitness Intelligence de una solución de gestión avanzada que le permitirá conocer en cada instante el estado de uso del gimnasio, el desempeño de cada usuario, las rutas que seguirán, etc. y además le lanzará alertas en tiempo real en caso de que algún usuario se encontrase realizando algún ejercicio incorrectamente, permitiéndole así actuar rápidamente y consiguiendo reducir las lesiones y aumentar la vida útil de sus aparatos.

Estamos convencidos de que las repercusiones de la disrupción que estamos introduciendo en el modelo de negocio tradicional del fitness es revolucionaria. Plasmamos nuestros objetivos y nuestra apuesta de futuro en nuestra **Misión** y nuestra **Visión**.

Nuestra **Misión**

Nuestra razón de ser es ofrecer soluciones de inteligencia de negocio avanzadas a centros de fitness para aumentar su productividad, su eficiencia y la fidelización de los usuarios.

Creamos tecnologías innovadoras y técnicas Big Data que permitan una transformación digital de la experiencia deportiva en tiempo real.

Apoyándonos en el conocimiento avanzado del consumidor, llevamos adelante una estrategia de diferenciación, optimización e innovación de servicios que nos permita convertirnos en líderes del mercado español.

Para ello, nos enfocamos en los consumidores finales, orientando nuestros servicios a la mejora del rendimiento deportivo, la salud y la experiencia del cliente, desarrollando el personal trainer digital para todos y cada uno.

Apostamos por un modelo de desarrollo sostenible, respetuoso con el medio ambiente y socialmente responsable que vele por la integridad y los derechos de las personas.

[PROBLEMA DE NEGOCIO]



Situación del sector fitness en España

El negocio del fitness es un sector en pleno auge, en Europa, ha crecido un 4% en 2015 generando 26,8 billones de euros y con un aumento del 9% en la afiliación este último año. En cuanto a España, la penetración de gimnasios y centros de fitness está por encima de la media europea (7,4%), un 10,2,% de la población española es socio de un centro de fitness.

Este sector ha generado 4.598 millones de euros en 2015 en España con un crecimien-

to medio anual del 5,1% entre 2011 y 2015. Además, según las previsiones realizadas en el informe de MarketLine sobre gimnasios, salud y centros de fitness de febrero de 2016, este sector generará 5.877 millones de euros en 2020, un 27,8% más que en 2015.

Se pueden distinguir 3 modelos diferenciados en función de los precios y servicios ofrecidos: premium, mid market y low cost. Los gimnasios premium son centros con cuotas altas (a partir de 60€) y gran cantidad de servicios como peluquería, cafetería, spa, nutrición y fisioterapia. Los gimnasios mid market son

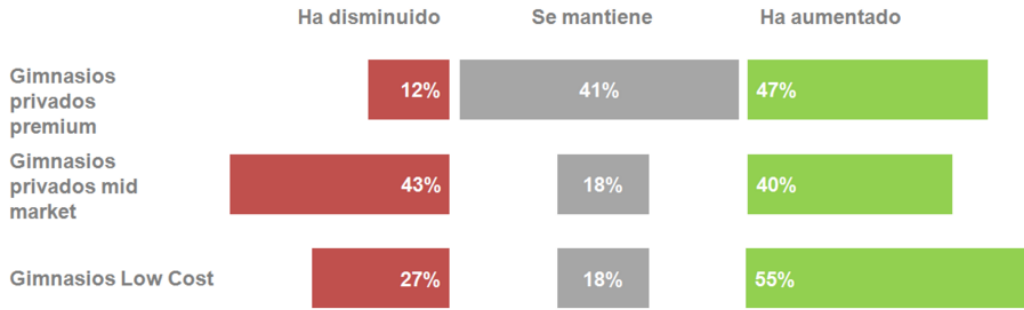
aquellos cuyos precios están en torno a los 40 euros y ofrecen menos servicios en comparación con los gimnasios premium. Finalmente, se denomina low cost a los centros cuyas cuotas son inferiores a 30 euros y que ofrecen servicios básicos.

Año	Millones €	% Crecimiento
2015	4.598,0	8,0%
2016	4.728,5	2,8%
2017	4.985,6	5,4%
2018	5.262,7	5,6%
2019	5.559,9	5,6%
2020	5.877,0	5,7%

Informe Market Line Gimnasios, Salud y Centros de Fitness. Febrero 2016

En cuanto a la facturación de estos centros, este último año casi todos los tipos de gimnasios han aumentado su facturación. La excepción es el gimnasio privado mid market, donde la facturación ha disminuido en un 43% de los centros según el Estudio Anual del Mercado del Fitness en España 2015 realizado por Life Fitness. Cabe destacar los centros low cost, en los que el 55% de los gimnasios han aumentado su facturación.

**FACTURACIÓN BRUTA
RESPECTO DEL AÑO
ANTERIOR**



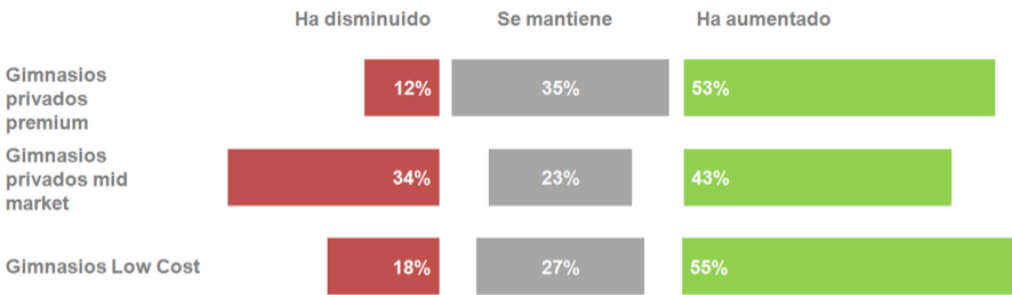
Estudio Anual del Mercado del Fitness en España 2015. Life Fitness

último año son los Estudios de Entrenamiento Personal y Gimnasios Low Cost, siendo los gimnasios mid-market y municipales los que más problemas tienen para retener a sus socios.

La irrupción del fenómeno de los gimnasios low cost ha generado una ruptura en el mercado, provocando que muchos operadores, especialmente los situados en el mid market, pierdan clientes debido a una baja diferenciación y estrategia de precios poco clara.

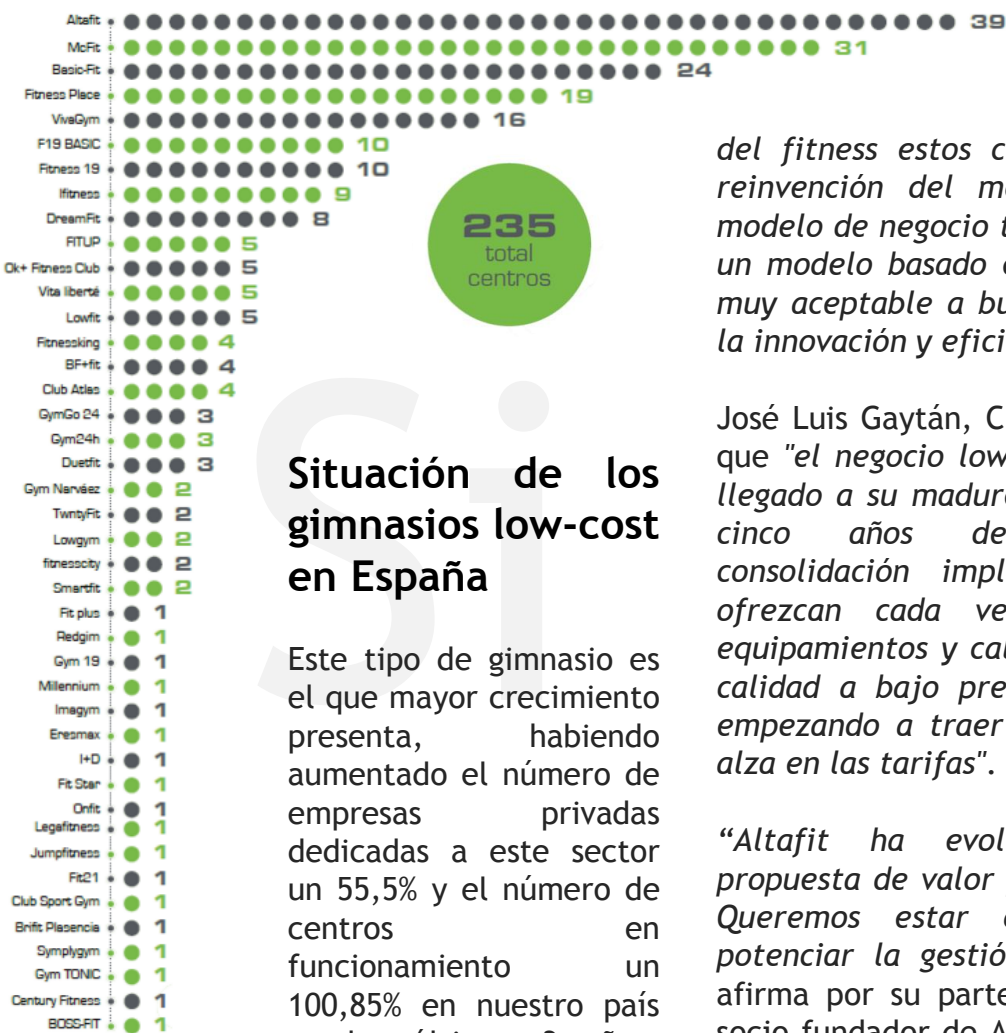
Respecto a número de clientes, los clubes que más aumentan en clientes activos este

**CLIENTES ACTIVOS
RESPECTO DEL AÑO
ANTERIOR**



Estudio Anual del Mercado del Fitness en España 2015. Life Fitness





Situación de los gimnasios low-cost en España

Este tipo de gimnasio es el que mayor crecimiento presenta, habiendo aumentado el número de empresas privadas dedicadas a este sector un 55,5% y el número de centros en funcionamiento un 100,85% en nuestro país en los últimos 3 años.

Debido a que este modelo requiere escalas mayores para obtener rentabilidad, tiende a la concentración copando 9 empresas el 70,6% de los centros. Es por esto que el número de centros crece más rápido que el número de empresas que los gestionan.

El modelo de negocio tradicional de estos gimnasios se basaba en cuotas bajas y el autoservicio por parte de sus usuarios, pero en la actualidad se encuentra en un proceso de cambio donde los precios bajos y el “self-service” ya no serán suficientes para captar y fidelizar clientes. Esta tendencia es corroborada por los principales Managers y CEOs que se exponen a continuación.

Según Marcos Verburgh, General Manager de Dreamfit, “debido a la sobreoferta existente, la recuperación del consumo y la asunción del clientes del nuevo precio

del fitness estos centros requerirán una reinversión del modelo de negocio. El modelo de negocio tiene que evolucionar a un modelo basado en ofrecer una calidad muy aceptable a buenos precios gracias a la innovación y eficiencia operativa”.

José Luis Gaytán, CEO de Fitness19 afirma que “el negocio low cost en el fitness, ha llegado a su madurez en 2015 después de cinco años de desarrollo. Esta consolidación implica que los centros ofrezcan cada vez mejores servicios, equipamientos y calidades ofreciendo alta calidad a bajo precio. Esta mejora está empezando a traer un pequeño ajuste al alza en las tarifas”.

“Altafit ha evolucionado hacia una propuesta de valor por encima del precio. Queremos estar cerca del cliente y potenciar la gestión de su experiencia”, afirma por su parte José Antonio Sevilla, socio fundador de Altafit, la empresa líder en gimnasios low cost.

“El modelo de negocio tiene que evolucionar a un modelo basado en ofrecer una calidad muy aceptable a buenos precios gracias a la innovación y eficiencia operativa”

-- Marcos Verburgh, Dreamfit

Fran Frago, Director General de Basic Fit Spain, corrobora esta nueva tendencia afirmando: “La cantidad de competidores que existen hacen que la calidad del servicio suba manteniendo el precio. Los clubes se construyen mejor, tenemos muchos más servicios complementarios al alcance de los socios y las redes sociales comienzan a marcar una diferencia”.

Estudio de mercado: Opinión de los usuarios de los centros low cost

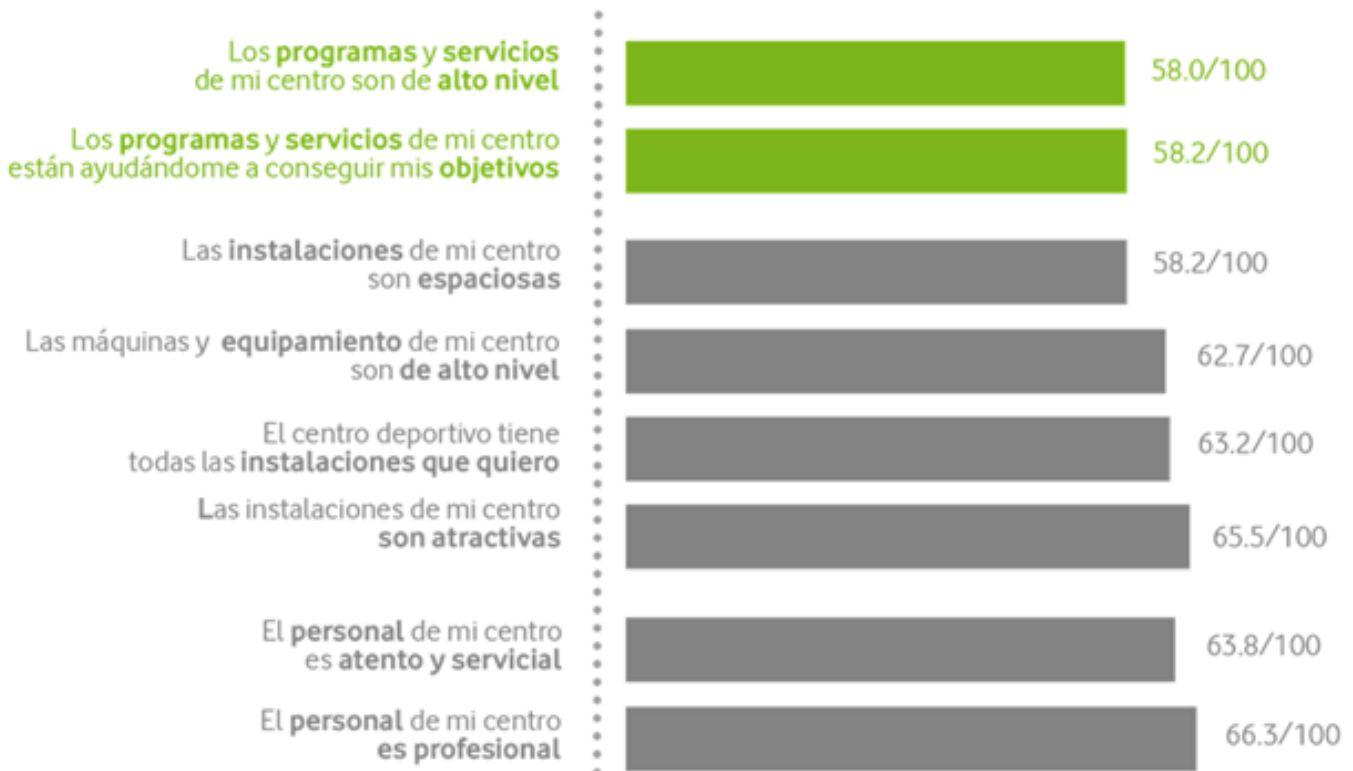
Gracias al cuestionario recogido en el 4º Informe de Gimnasios Low Cost en España de enero 2016 realizado por Valgo donde se toma como muestra a 15.820 personas usuarias de los gimnasios low cost (8.462 mujeres y 7.358 hombres), hemos podido estudiar el perfil y la percepción del cliente en centros de fitness españoles low cost.

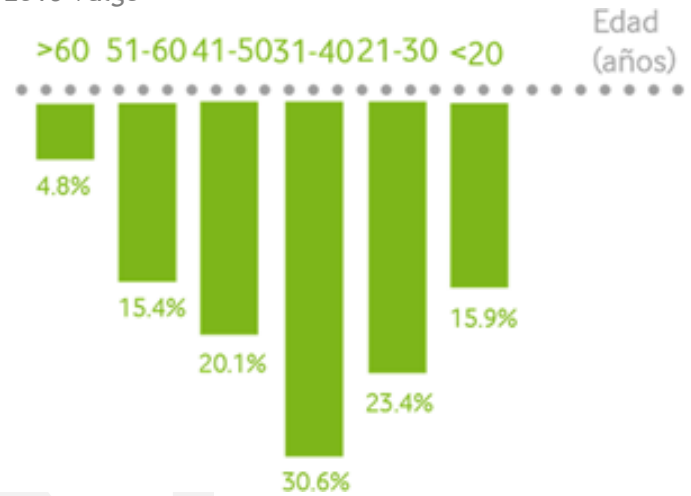
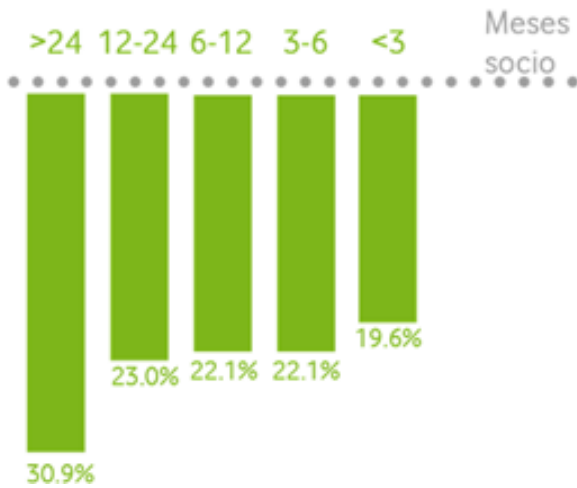
Cuando se preguntó el número de veces que contactaba el personal con el cliente cada vez que asistía, el 21% afirmó que nunca, una vez el 29,8%, y el 22,7% más de 3 veces por asistencia. Asimismo, el 60,2% afirmó que el entrenamiento se lo planificaba el propio cliente, frente al 18,6% que lo hacía el técnico deportivo. En concreto, solo el 3,1% llevaba el entrenamiento mediante una aplicación informática.

Referente a los hallazgos encontrados sobre la percepción de diferentes factores sobre el centro de fitness, indicar que la puntuación más baja fue para los programas y servicios, teniendo en cuenta que el mínimo de puntuación fue un 1 y el máximo un 10. En estos casos, los programas y servicios se puntuaron con un 5,80 y 5,82.

En relación al Net Promoter Score, indicador para medir la lealtad del cliente se basa en una sola pregunta: “¿Qué probabilidad hay de que recomiende su gimnasio a un familiar o amigo?”. Para ello se les pide calificar en una escala de 0 a 10, donde 0 es muy improbable y 10 es definitivamente lo recomendaría. Según los resultados, los clientes se clasifican en promotores, pasivos y detractores:

- Los que responden asignando 9 o 10 puntos: **Promotores**
- Los que asignan 7 u 8 puntos: **Pasivos**
- Los que otorgan 6 puntos o menos: **Detractores**





Para obtener un resultado, se restan los detractores a los promotores y se consigue un porcentaje de promotores para medir la calidad del servicio.

Cabe destacar que este indicador en todos los casos se obtiene un indicador negativo, existen más detractores que promotores, siendo los más positivos en los clientes de más de 60 años y en los que llevan inscritos menos de 3 meses. Por lo tanto, podemos concluir una baja lealtad o satisfacción por parte de los usuarios hacia el gimnasio al que asisten.

Necesidad que vamos a cubrir

Gracias al estudio explicado anteriormente, hemos podido identificar el problema de negocio de los gimnasios low cost.

Debido a la sobreoferta existente, la baja satisfacción en relación a los programas y servicios ofrecidos, la demanda por parte de los usuarios de una mayor personalización y calidad y un Net Promoter Score negativo, identificamos el problema de negocio de los gimnasios low cost: la necesidad de retención de los usuarios y diferenciación por parte de los centros mediante una mejora de la calidad de los programas y servicios.

Fitness Intelligence permitirá que estos centros puedan diferenciarse de la competencia al ofrecer una solución única, que permite una personalización total de los entrenamientos y democratización del servicio de entrenador personal, suponiendo una mejora en la satisfacción del cliente y del Net Promoter Score y con ello acrecentando la captación y retención usuarios. Además, gracias a nuestra solución, la gestión del gimnasio será más sencilla y en mayor detalle, capaz de detectar la senda de abandono de los usuarios.



Gimnasios Low Cost

Para llevar a cabo nuestro proyecto, hemos investigado los nombres más sonantes entre las cadenas low-cost.

Hemos hablado con ellos y les hemos hecho preguntas determinantes para asegurarnos de que nuestro proyecto le será de gran utilidad.

Ninguna de estas cadenas tiene una aplicación que haga nada similar a lo que nosotros ofrecemos a nuestros clientes. Dentro de estas cadenas podemos destacar la aplicación que tiene la cadena Fit Up, la cual es útil para que los clientes puedan hacer un seguimiento de sus entrenamientos, pero la mayoría de los datos los tienen que introducir los usuarios manualmente, lo cual genera que el olvidarse un día el teléfono a la hora de entrenar puede hacer que se olviden de meter algún ejercicio que hayan hecho y de cómo lo hayan hecho.

Por otro lado, tampoco hay ninguna cadena que controle mediante una APP a la que puedan acceder los usuarios, la gente que hay en sala, ni la ruta adecuada que los usuarios deben seguir para optimizar los recursos de las instalaciones. Este es otro de los grandes problemas a los que se enfrentan estas cadenas: a las horas punta, tienen determinadas máquinas de ejercicios totalmente colapsadas, sin embargo hemos podido observar que otras no se utilizan constantemente.

Algunos de estos centros cuentan con el servicio de personal trainer y hemos visto que el precio medio es de 20 euros por hora, lo cual es un precio competitivo pero la disponibilidad no es total, lo tienen que solicitar y según los foros visitados, los usuarios se quejan de que en horas punta hay lista de esperar para tener al entrenador a tu disposición.

Entendemos que un personal trainer puede ser de gran ayuda a la hora de entrenar, pero ninguno de ellos será capaz de medir con la precisión que medimos nosotros el entrenamiento de los usuarios. Por otro lado, sin nuestra tecnología el personal trainer tampoco podrá hacer un estudio detallado sobre el histórico de entrenamientos del usuario, su única herramienta para ayudarle será escuchar lo que el usuario dice, de manera que entendemos que en la mayoría de los casos una sesión no será tiempo suficiente para ayudarle correctamente.

Como se puede observar en la tabla, las extensiones medias de las instalaciones de estas cadenas de gimnasios supera los 1500 metros. Tras estudiar cada una de estas cadenas nos hemos dado cuenta de que se rigen más o menos por el mismo patrón. El personal con el que cuentan es insuficiente, en algunos casos hay una persona en recepción y otra dando clases de actividades y cada hora se van rotando, de manera que sólo hay una persona dando las clases y otra en recepción, dejando la sala desatendida. Cuando el usuario necesita algo, tiene que ir a recepción a preguntar y en la mayoría de los casos la persona que está atendiendo en recepción está ocupada con otros asuntos.

Debido a estos dos puntos (extensión de los gimnasios y carencia de personal) y teniendo en cuenta que la media de clientes de estos gimnasios es de 2700 personas, entendemos que nuestra aplicación les será más que útil para poder mejorar la atención al usuario ya que en la mayoría de los casos será nuestra aplicación mediante notificaciones push la que informe a nuestros usuarios de que están haciendo algo mal de manera que la carencia de tener el número adecuado de monitores en sala será menos perceptible para los clientes de los gimnasios.

Gimnasio	Precio/ mes	Extens. (m2)
Dreamfit	22,9	3000
Fitness 19	19,9	1500
FitUp	22	2000
lfitness	23	3000
Basic Fit	19,9	2000

Datos recogidos a partir de investigación propia

Cuando el usuario pregunte al monitor, este tendrá la herramienta adecuada para poderle guiar de una manera mucho más precisa

¿Qué ofrece la competencia?

Hemos identificado a la empresa **Technogym** como competencia. Utilizan una herramienta llamada **PRESCRIBE**, la cual permite que los entrenadores personales puedan crear una variedad de programas de entrenamiento personalizados para los clientes, administrables y accesibles desde la APP.

Gracias a ella, el personal puede crear y organizar una biblioteca de hasta 2.000 programas adaptables a las necesidades de los clientes. Son planes ya diseñados y buscan cual se adapta más al cliente, no es un plan diseñado específicamente para el cliente.

Mediante la APP pueden recuperar toda la información sobre los clientes rápidamente.

Por otra parte, sus clientes pueden empezar a hacer ejercicio de forma inmediata con un solo clic, dado que el entrenamiento asignado se transfiere al equipo con solo iniciar sesión.

Sus clientes pueden consultar y gestionar sus programas de entrenamiento desde cualquier lugar a través de su APP.

Además, los clientes pueden ver sus vídeos de ejercicios pregrabados en la aplicación móvil o a través de Internet en mywellness.com. Esta compañía fabrica la maquinaria, y en muchas de estas máquinas incluyen pantallas similares a las de la figura 1.

Como se puede observar en la imagen, da datos de las calorías gastadas, pulsaciones, tiempo, etc. Ninguno de estos datos, salvo las pulsaciones, son datos personalizados. Para que la máquina pueda ver las pulsaciones el usuario tendrá que apoyar la manos sobre una placa metálica que hay en los agarradores del aparato y el resto de datos que ofrece son genéricos, salvo que el usuario cada vez que empiece un ejercicio vaya rellenando manualmente sus datos.



Fig. 1: Interfaz de usuario de PRESCRIBE

¿Qué nos diferencia de ellos?

Nuestra tecnología está basada en tiempo real, no nos vamos a limitar a crear una APP con rutas donde los usuarios de los gimnasios puedan ver como se hace un ejercicio, para el desarrollo de esta labor, se pueden poner video tutoriales en las salas dedicadas a las actividades colectivas. Tampoco nos vamos a limitar a llevar a la APP la información que se puede ver en cualquier máquina de cardio, creemos que es más sencillo que el usuario vea esta información en la propia máquina de correr o bici, las cuales vienen siempre dotadas de una pantalla informativa en la que se da esta información.



Fitness Intelligence: Cien pasos más allá

Nuestro objetivo es llegar donde nadie llega, monitorizar a tiempo real absolutamente todos los ejercicios que se pueden hacer en una sala de fitness, y será eso lo que nos diferencia de la competencia, podremos darle al usuario una ruta perfectamente adecuada a sus necesidades y lo que es más importante, le daremos a los monitores y al propio usuario una herramienta que corregirá cualquier ejercicio que no esté realizando correctamente.

Además le daremos al usuario una ruta que se adecúe a sus necesidades y además a las circunstancias del centro, con la finalidad de evitar la congestión de determinadas máquinas, de manera que todo usuario podrá terminar con su rutina en el menor tiempo posible, lo cual posibilitará que nuestro cliente pueda dar cabida a más usuarios sin modificar su infraestructura al tiempo que el usuario está más contento al no tener que esperar colas.

Gracias a nuestra tecnología de última generación, la APP informará mediante alertas push de cualquier error en el entrenamiento de un usuario. Esta alerta

saltará en la APP con un pitido y una vibración constante hasta que el cliente acceda a la aplicación, de manera que aunque los monitores no puedan atender al usuario en ese preciso momento, nos aseguraremos de que nadie sufra una lesión por una falta puntual de personal.

Los usuarios encontrarán en su app la ruta a seguir, la intensidad del entrenamiento que están realizando, los ejercicios que le quedan para terminar y la gente que hay en sala. Con este último dato, los usuarios podrán decidir con antelación si prefieren ir a otra hora para evitar aglomeraciones.

Por otro lado, en la parte derecha del panel, encontrarán las alertas, con las que sabrá si hay algo que no está haciendo bien. Esta alerta llegará tanto al usuario como al monitor de sala, de manera que esas conductas erróneas se podrán corregir rápidamente.



Fig.2: Interfaz de usuario de Fitness Intelligence

Propuesta de valor

Hacia el cliente (gimnasio):

- **Ampliar inteligencia de negocio:** Combinando de tecnología, herramientas y procesos les permitiremos transformar los datos que almacenen en información para el desarrollo y control de su negocio.
- **Reducción de abandonos:** Usando como base los patrones de conducta de otro cliente que anteriormente se hayan dado de baja, podremos predecir cuales son los clientes de alto riesgo de baja y así nuestro cliente podrá anticiparse para retenerlos
- **Optimización de la distribución de máquinas:** Elaboraremos rutas que optimizarán los entrenamientos de los usuarios de los clientes y las infraestructura de los gimnasios
- **Reducción de colas:** Mediante el diseño de rutas optimas de entrenamiento, la máquinaria de los gimnasios tendrá menos demora para que pueda utilizarla el usuario
- **Clustering de clientes:** Nuestros clientes conocerán los gustos y las costumbres de los usuarios, de esta manera podrán ofrecerles rutas personalizadas
- **Detección de anomalías en el uso:** Podremos predecir por patrones anteriores que máquina va a estropearse
- **Optimización de recursos (RRHH):** El personal del gimnasio tendrá una herramienta con la que podrá controlar a muchos más usuarios en menos tiempo y mucho mejor que anteriormente
- **Eficiencia energética:** Al estar localizados los usuarios dentro del gimnasio podrán regular la energía a consumir en determinadas áreas del gimnasio (luz, ventilación, calefacción)

Hacia el consumidor / usuario (e indirectamente hacia el cliente):

- **Plan de entrenamiento personalizado digitalizado:** Tendrán una ruta personalizada al 100% que se modificará automáticamente dependiendo de las necesidades de cada usuario.
- **Seguimiento de la actividad física:** Podrán hacer un seguimiento de sus rutinas de entrenamiento en su app
- **Reducción de probabilidad de lesiones:** En caso de que estén en riesgo de lesión porque no estén entrenando correctamente, serán avisados.
- **Gamificación:** Incorporando estrategias a la hora del entrenamiento que pueden apoyar y motivar al usuario a realizar de una manera entretenida los ejercicios.
- **Democratizar acceso a personal trainer:** A partir de ahora gracias a nuestra tecnología todo el mundo podrá disfrutar de unos servicios similares a los de un personal trainer
- **Reducción de tiempo de espera en uso de máquinas:** Mediante la elaboración de rutas optimizaremos el uso de la infraestructura del gimnasio.
- **Atención del gimnasio más inmediata y personalizada:** El monitor de sala podrá darles una atención eficaz e inmediata ante cualquier anomalía en su entrenamiento.
- **Conocimiento en tiempo real de la ocupación del centro:** Sabrán en cualquier momento cuanta gente hay en el centro para poder ir a la hora que más les convenga
- **Optimización del tiempo en el gimnasio:** Con estas rutinas optimizadas, sacarán más rendimiento a su entrenamiento en menos tiempo



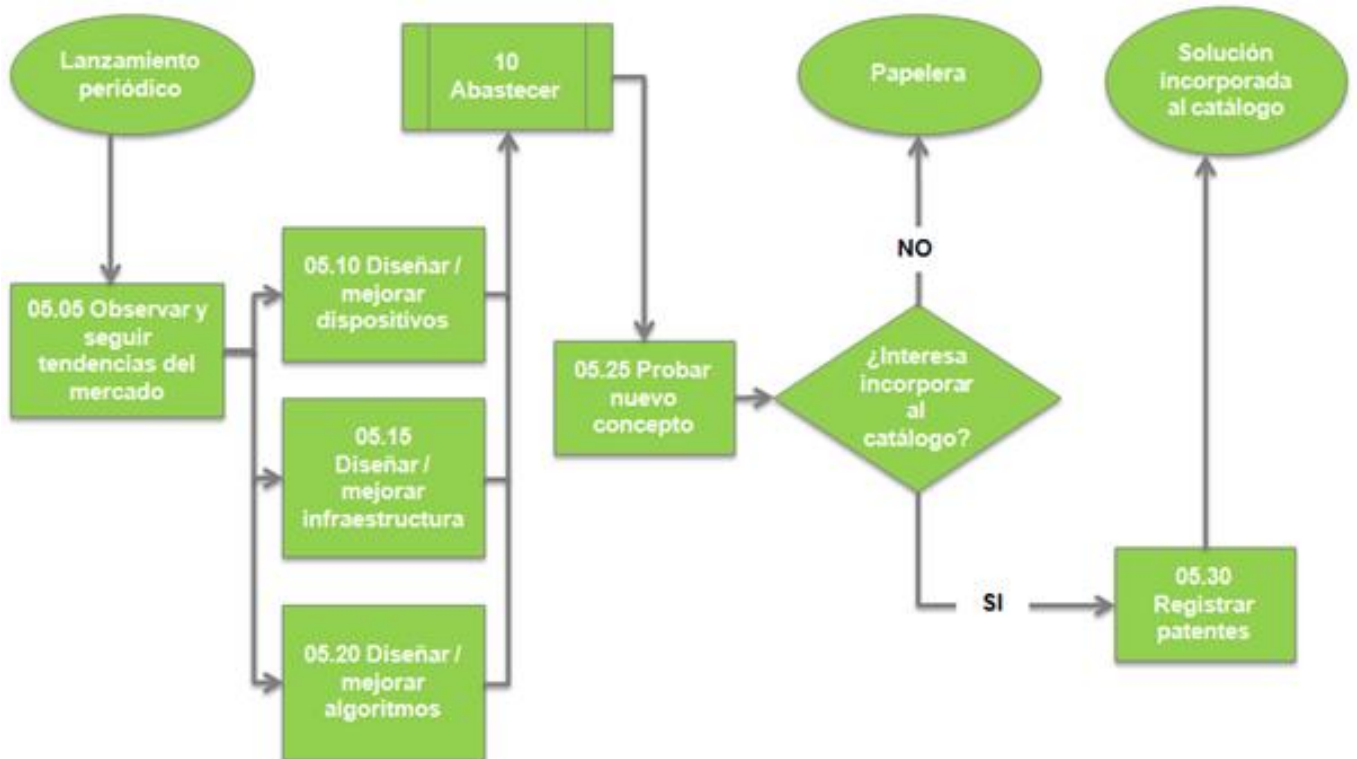
Cadena de valor de Fitness Intelligence. Mayo 2016

Procesos

Para describir el desarrollo de las actividades de Fitness Intelligence con el fin de crear valor a nuestros clientes finales, hemos construido una cadena de valor compuesta por 7 procesos, de los cuales hemos desarrollado 6.

El primero de ellos, *Diseñar y mejorar algoritmos y tecnología*, contempla el diseño y mejora de los dispositivos, infraestructura, algoritmos de machine learning además de realizar un seguimiento del mercado junto con los avances tecnológicos y registro de patentes.

Periódicamente se observará y seguirá las tendencias del mercado, que servirán para diseñar y mejorar los dispositivos, la infraestructura y los algoritmos para que más adelante pasen al proceso de abastecimiento donde se realizará una prueba de concepto. En función de los resultados de esta prueba, se decidirá si interesa incorporar la innovación o mejora al catálogo de productos y servicios. En caso afirmativo, se registrarán las patentes relativas a esta solución o innovación y se incorporará al catálogo.

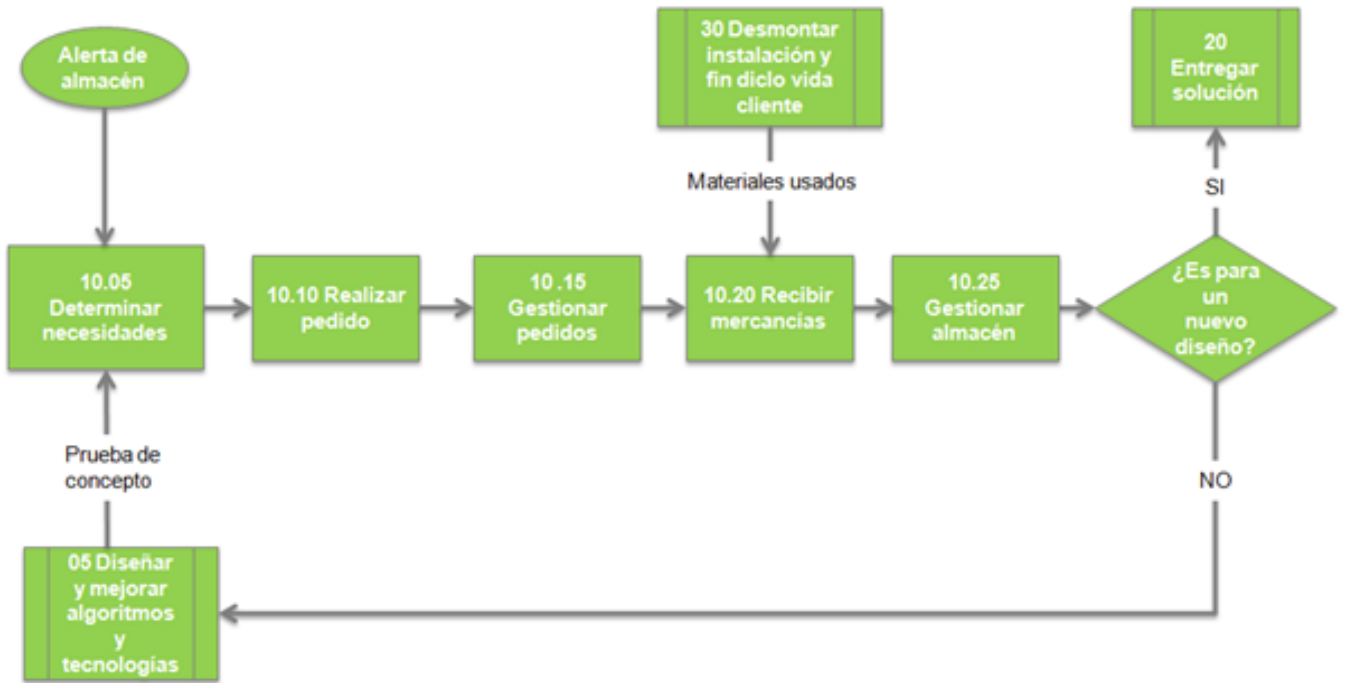


Proceso 05 - Diseñar y mejorar algoritmos y tecnología. Mayo 2016

El segundo proceso es el abastecimiento, donde se determinará las necesidades, se realizarán y gestionarán pedidos, se recibirán mercancías y se gestionará el almacén.

En este proceso en primer lugar se determinan las necesidades a satisfacer, ya sean por falta de inventario en el almacén o de mejora y diseño de algoritmos y tecnologías. Una vez definidas, se realizará

el pedido el cual se gestionará hasta recibir las mercancías. Fitness Intelligence las recibirá mediante dos vías, a través de la recepción de nuevos pedidos o reutilizando materiales usados una vez finalizado el ciclo de vida del cliente. Estas mercancías se gestionarán en el almacén para más adelante entregar la solución en el caso de que sea para un nuevo diseño o servirán para el proceso de diseño y mejora de algoritmos.



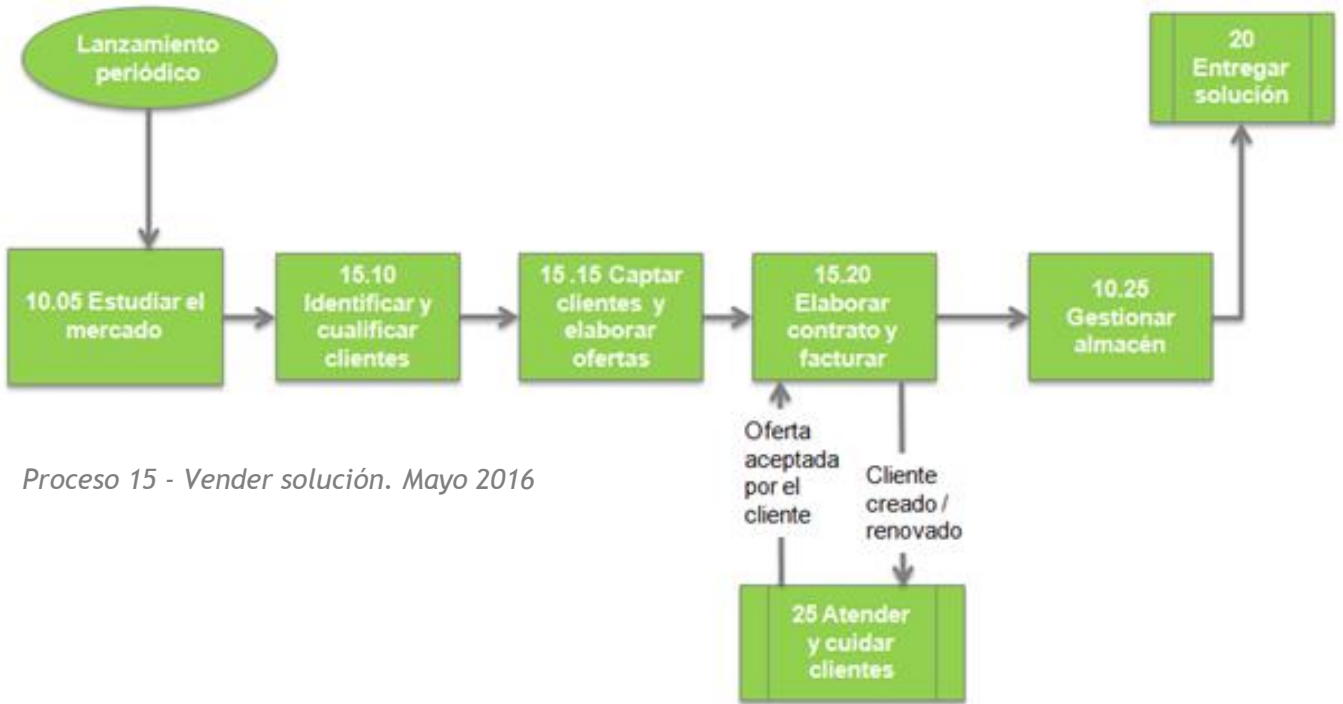
Proceso 10 - Abastecer. Mayo 2016



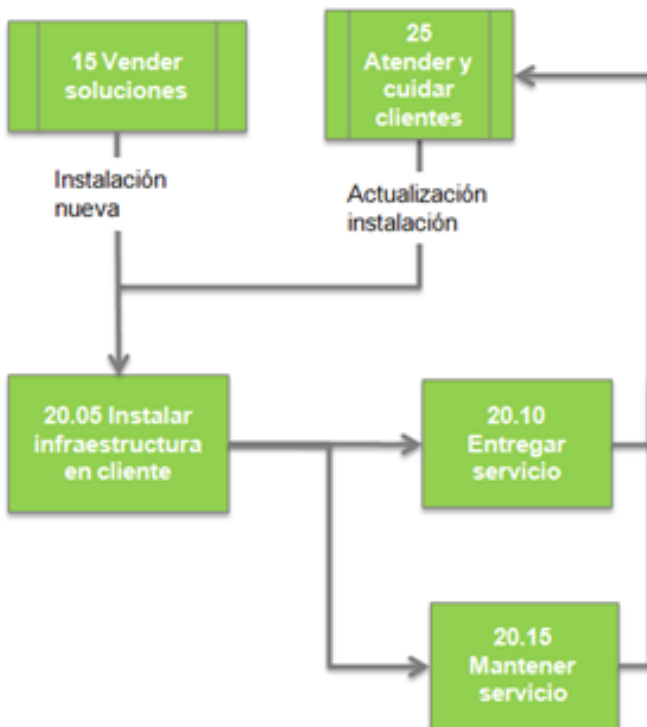
El tercer proceso abarca las actividades que desarrollamos durante la venta de nuestras soluciones.

Se estudiará el mercado para así poder identificar y cualificar a los clientes, captarlos y

elaborar ofertas efectivas. Una vez el cliente acepta la oferta o renueva se procederá a la firma del contrato y facturación y se gestionará en almacén los componentes necesarios para realizar la entrega de la solución.

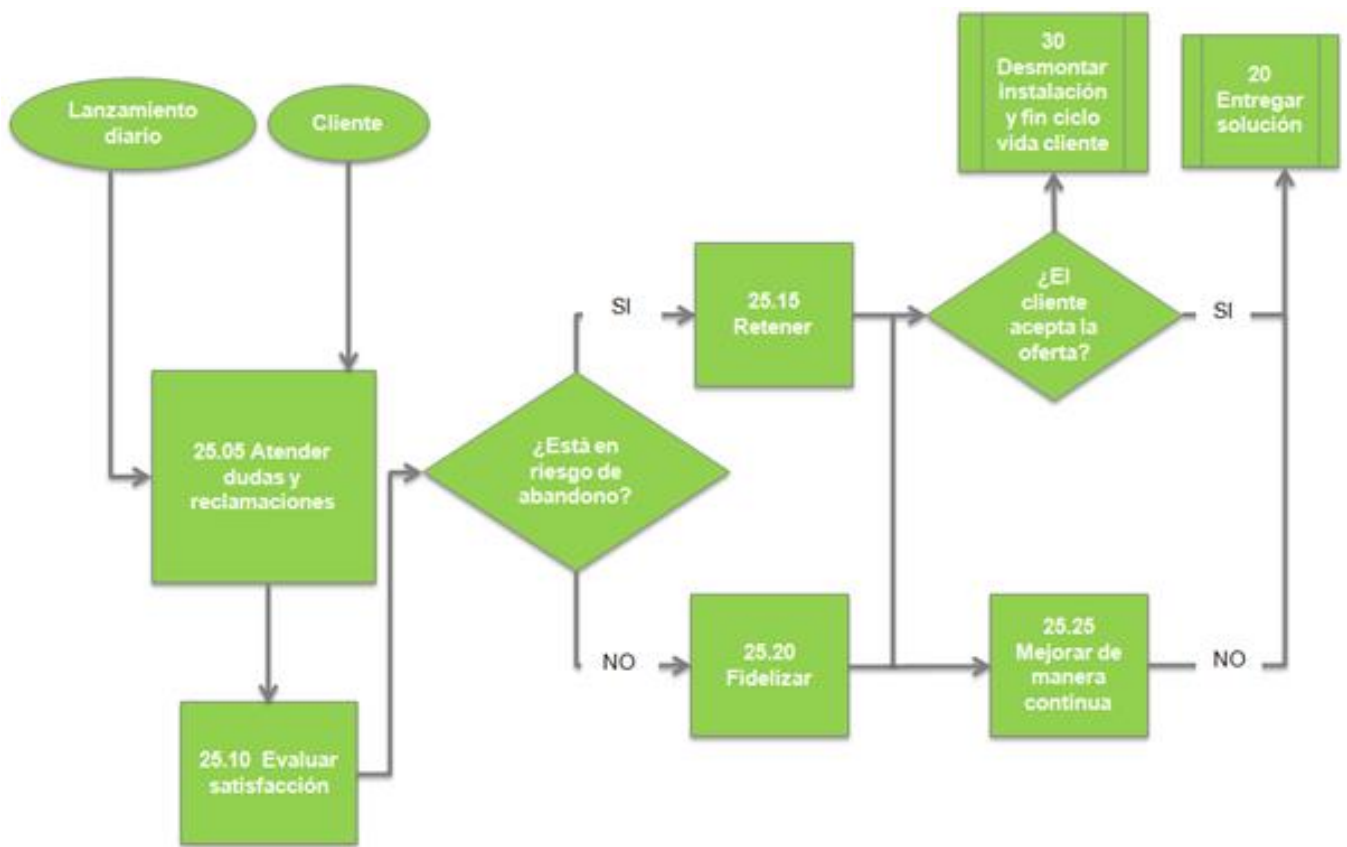


Proceso 15 - Vender solución. Mayo 2016



Proceso 20 - Entregar solución. Mayo 2016

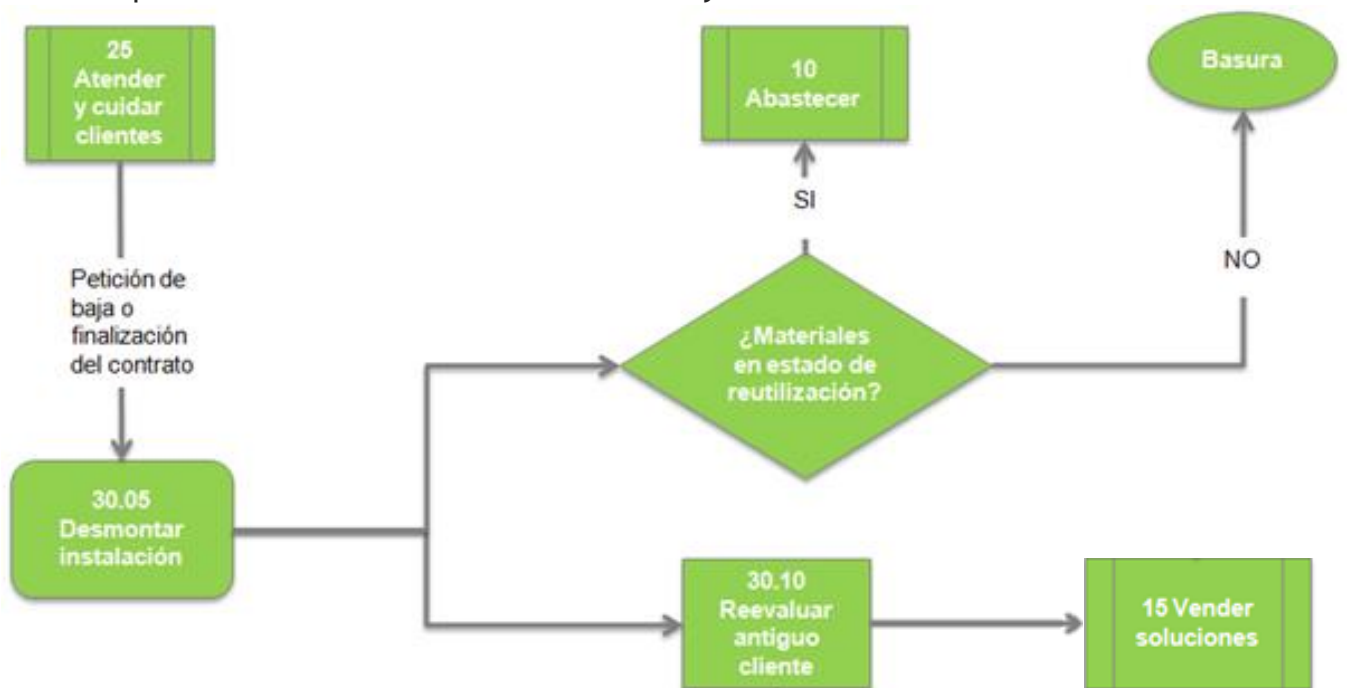
El proceso de entrega de la solución consiste en primer lugar en instalar la infraestructura en los gimnasios ya sea debido a una nueva contratación o a una actualización de la instalación. Una vez esté en funcionamiento, se entregará y mantendrá el servicio para así atender y cuidar de nuestros de clientes.



Proceso 25 - Atender y cuidar al cliente. Mayo 2016

Para el proceso de atender y cuidar al cliente, Fitness Intelligence atenderá sus dudas y reclamaciones para evaluar la satisfacción y con ello establecer si el cliente está en riesgo de abandono. En el caso de que se determine que el cliente pueda prescindir de nuestros servicios, se intentará retenerlo ofreciendo una nueva oferta personalizada en función de sus

necesidades. Aunque se evalúe que el cliente no esté en riesgo de abandono, se realizarán acciones de fidelización para mejorar de manera continua su satisfacción. En la situación de que se ofrezca una oferta y que sea aceptada se procederá a la entrega de la solución y en caso contrario se desmontará la instalación y finalizará el ciclo de vida del cliente.



Proceso 30 - Terminar ciclo de vida del cliente. Mayo 2016

En el proceso de finalizar el ciclo de vida del cliente se desmontará la instalación y en el caso de que los materiales se puedan reutilizar pasarán al proceso de abastecimiento utilizándose para posteriores instalaciones. Además, se reevaluará al antiguo cliente que ha decidido darse de baja para en un futuro volver a vender la solución.

Para dar soporte al negocio, Fitness Intelligence realizará las actividades de gestión de recursos humanos, formará a sus empleados, mantendrá sus servidores y equipos, gestionará finanzas y contabilidad, planificará, supervisará y coordinará las actividades del negocio.

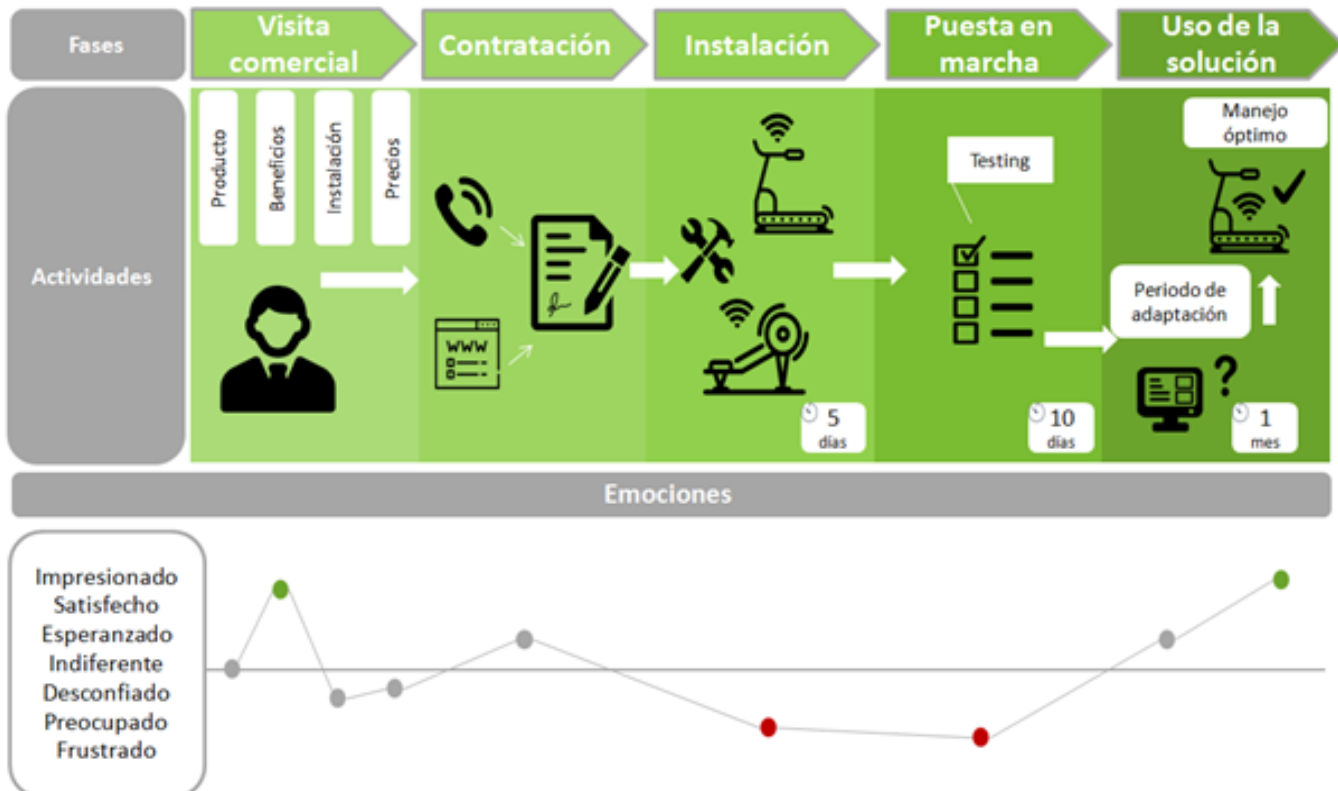
cial, contratación, instalación, puesta en marcha y, por último, el uso de la solución

Una vez se ha estudiado y clasificado a los clientes potenciales, un agente comercial los visitará y explicará las características y los beneficios de nuestros productos, cómo es el proceso de instalación y los precios de la solución. Identificamos que en esta fase, el cliente puede sentirse desconfiado debido al proceso de instalación que a priori puede parecer engorroso. Por ello, es vital que el comercial inspire confianza mostrando a nuestro posible cliente las garantías de nuestros servicios, destacando que la instalación tendrá lugar en por las noches o cuando el centro esté cerrado para no interferir en ningún momento en su funcionamiento.

En el caso de que el cliente quiera contratar nuestros servicios, podrá ponerse en contacto con nuestra central o con su agente por vía telefónica o a través de nuestra página web gracias a un botón click to call. En esta fase, será vital el correcto funcionamiento de la web y la respuesta rápida por parte del

Customer Journey

Para entender la experiencia de nuestros clientes, hemos elaborado un mapa de la experiencia del cliente desde la visita comercial al uso de la solución, donde distinguimos cinco fases: la visita comer-



departamento comercial, para así inspirar seguridad y seriedad a nuestro futuro cliente.

La fase de instalación de la solución tendrá una duración de 5 días y es seguida por la de puesta en marcha donde se realizará un testing durante 10 días para asegurarnos de su correcto funcionamiento. Estas dos fases son críticas para nosotros ya que el cliente se sentirá preocupado e incluso frustrado al ver que se están manipulando sus máquinas y haber realizado ya un pago inicial sin comenzar a ver resultados. Por ello, para tranquilizar y mostrar transparencia hacia el cliente, se hará partícipe al cliente mostrando todos los resultados y avances.

Una vez la solución funcione correctamente y haya pasado todos nuestros tests, el cliente podrá hacer uso de ella. Durante el primer mes, el personal

del gimnasio se encontrará en un periodo de adaptación donde comenzará a ver resultados pero se encontrará ante una solución única e innovadora que desconoce, sintiéndose esperanzado pero a la vez preocupado. Por este motivo, se realizarán cursos de formación y visitas de seguimiento asegurando la total satisfacción de nuestros clientes.

Para Fitness Intelligence es prioritaria la satisfacción del cliente, por ello, identificamos los touchpoints explicados anteriormente los cuales son críticos para garantizar nuestros valores de transparencia, confianza y cercanía.

Además, hemos realizado una primera aproximación al cuadro de mando estratégico que regirá las acciones de la compañía, estructurado según el modelo de Norton y Kaplan de Cuadro de Mando Integral:



Cuadro de Mando Integral a alto nivel



Introducción

El concepto de Fitness Intelligence se basa en una extensa monitorización del usuario y en la recogida de datos. Para ello es necesario la instalación de una infraestructura de sensorería potente en las instalaciones del cliente. Ello nos permitirá recopilar datos de forma autónoma, centralizarlos, y posteriormente realizar un análisis sobre los mismos. La instalación no es demasiado compleja, si bien utiliza sensores de desarrollo propio diseñados para insertarse perfectamente en las máquinas de ejercicio y facilitando laEl diseño de la solución incorpora varios elementos que en su conjunto permiten obtener varios tipos de datos de los usuarios para hacer un perfil completo de la actividad deportiva que realizan en los gimnasios. Un conjunto de sensores diversos y adaptados posibilitan este hecho. recogida de datos.

Los usuarios tendrán una aplicación móvil para ver su progreso y configurar sus rutinas de entrenamiento, además de otras funcionalidades como participar en retos de gamificación, obtener información sobre la ocupación del gimnasio y poder contratar servicios de nutricionistas y de médicos deportivos, entre otras cosas.

Dispositivos que instalaremos en los gimnasios

- Sensores de maquinaria: 1 por cada máquina. Incorporan un lector de RFID, un sensor para caracterizar el uso de la máquina y un módulo de comunicación Wifi.
- Pulseras: 1 por cada usuario. Monitorizarán el ritmo cardiaco y la posición del usuario. Contarán con un módulo de comunicación Bluetooth.
- Estación de comunicación Bluetooth & WiFi: recopilará los datos de la pulsera, y los envía por Internet a nuestros servidores.

Especificación de los dispositivos

Los sensores tendrán un microcontrolador integrado, con un módulo WiFi para realizar un conexión directa al servidor. El microcontrolador se programará para configurarlo con las conexiones de nuestro servidor (cifrado, protocolo, etc). El microcontrolador enviará la información por WiFi (protocolo TCP/IP) con una capa



por encima de codificación de datos. Con este sensor se puede comunicar a cualquier programa/software que esté en la nube y que conozca el protocolo. Lo cual quiere decir que se abrirá y empezarán a llegar los datos por ese canal, que luego serán traducidos en información aplicando el protocolo en cuestión.

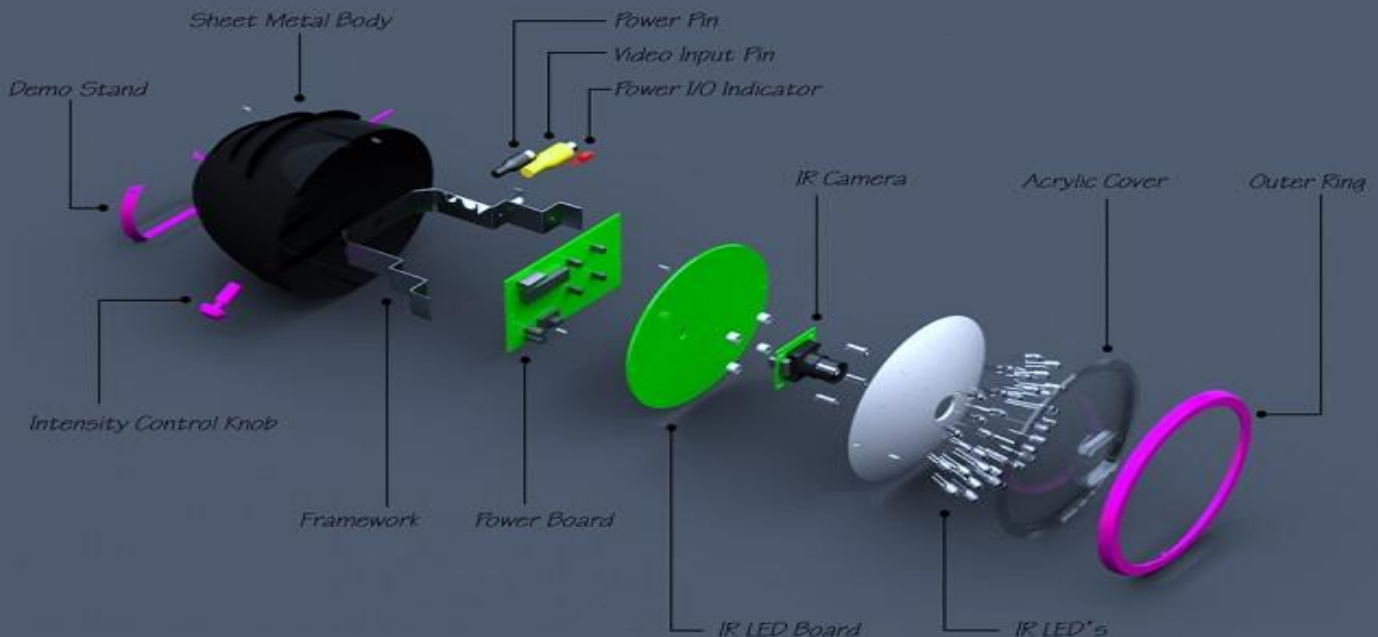
Las pulseras a su vez también contarán con un microcontrolador que se encargue de recopilar los datos de las lecturas de frecuencia cardíaca y los enviará por Bluetooth a una estación de comunicación, que se encargará de enviar dichos datos al servidor de Kafka. La etiqueta RFID pasiva también estará incorporada y se comunicará con el lector de RFID de cada máquina de forma independiente.

Para las pulseras y los sensores, gracias a la altísima experiencia y calidad del equipo técnico, serán desarrollados en exclusiva por Fitness Intelligence. De esta forma se tiene un conocimiento y control absoluto sobre esta parte crítica de la infraestructura y se innovará continuamente para mejorar el rendimiento, el consumo y el funcionamiento de los mismos. De esta manera también se podrá bajar los costes de producción, externalizando la manufactura a un país como China, lo cual aumenta considerablemente los márgenes económicos asociados.

Las siguientes partes son comunes tanto a los sensores de la maquinaria como a las pulseras de los usuarios: 1. Microcontrolador 2. Módulo Wifi/Bluetooth 3. Sensor 4. Batería 5. Carcasa.

La instalación: Punto de vista del usuario

Al entrar en el gimnasio, los usuarios recibirán una pulsera equipada con un lector de frecuencia cardíaca (pulsómetro), una etiqueta RFID pasiva y con conectividad Bluetooth.



Esta pulsera estará ligada a su perfil, para asociar correctamente los datos que se generan durante el entrenamiento. Con este paso comienza la experiencia del usuario, y ya se empieza a recopilar datos vitales del mismo.

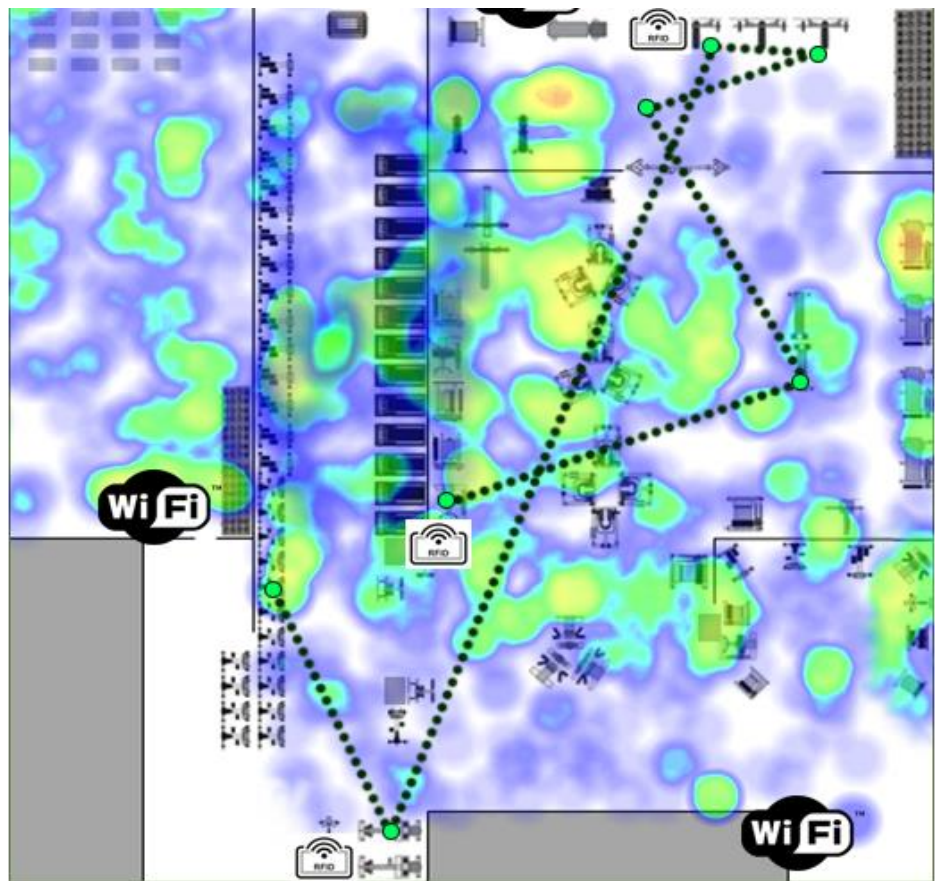
El usuario se dispone a realizar su rutina de entrenamiento. Las máquinas están equipadas de lectores de RFID y de sensores que miden el desempeño del usuario al realizar el ejercicio. Al sentarse para empezar el ejercicio, el lector RFID de la máquina identifica al usuario. En este momento, el usuario comienza su rutina de ejercicio, y los sensores de la máquina empiezan a recoger y enviar los datos de uso. La información generada por los sensores se envía a los servidores por WiFi. El sistema sabe qué usuario en concreto está realizando el ejercicio ya que se le ha identificado previamente por RFID.

El usuario ha completado su rutina. Una vez acabado el ejercicio, el usuario abandona el puesto de ejercicio, lo cual es detectado por el lector RFID al dejar de registrar la señal de identificación del usuario.

El usuario podrá consultar en su aplicación su desempeño, comparando los resultados del entrenamiento con su historial y la aplicación podrá realizar recomendaciones y sugerir la siguiente actividad. Por ejemplo si ha detectado que el desempeño del usuario en un grupo de músculos en

particular ha sido más bajo que el desempeño de las últimas sesiones, la aplicación podrá recomendar al usuario que deje de ejercitar esa área para reducir el riesgo de sobrecarga/lesión. Y le sugerirá otro ejercicio distinto para trabajar otra área.

El supervisor de la instalación contará con un software que le permita de un vistazo ver la ocupación de la sala, que máquinas se están utilizando y cómo. Contará con un sistema de alertas que le informará del mal uso de las máquinas para que pueda intervenir y corregir el comportamiento del usuario. También



Panel de Control del Gimnasio de Fitness Intelligence

podrá consultar los perfiles de los usuarios en tiempo real, para poder supervisar mejor la rutina e interceder si detecta un comportamiento anómalo.. Además contará con otras funcionalidades como la previsión de roturas de las máquinas e informes generales sobre la actividad.

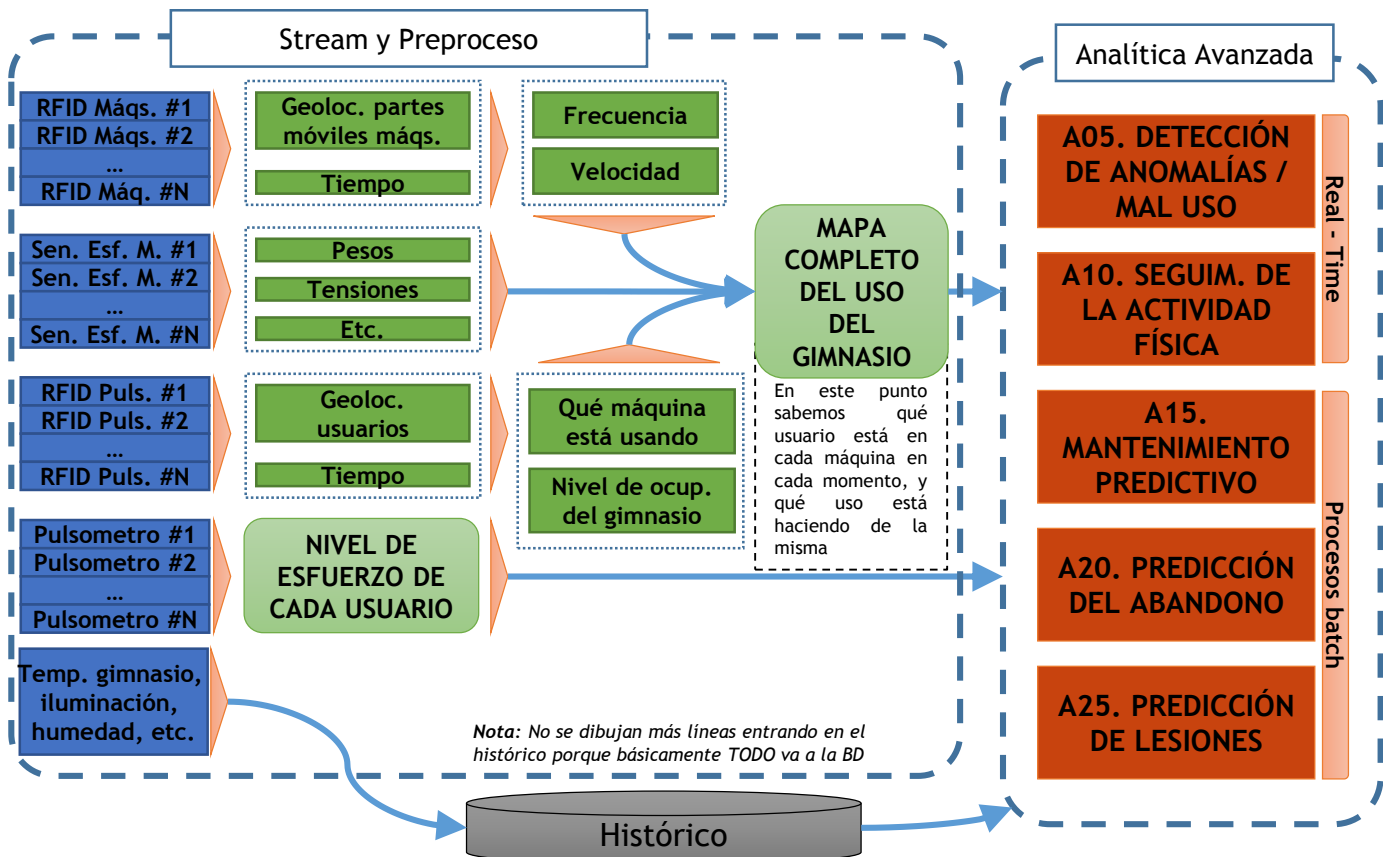
[DATOS ESTRUCTURADOS]



A partir de los datos generados por los sensores de las máquinas, de las pulseras y de los sensores ambientales del gimnasio, realizamos análisis tanto en tiempo real como batch para desplegar toda la oferta de valor.

Durante la 1ª fase de implantación, los algoritmos que desplegaremos serán:

- **Detección de anomalías** - Detección de patrones de uso indebido de las máquinas (TR)
- **Seguimiento de la actividad física** - Decisión en tiempo real del próximo ejercicio en función del performance actual y pasado, objetivos del usuario y niveles de ocupación (TR)
- **Mantenimiento predictivo** - Detección profunda de factores críticos del funcionamiento (batch)
- **Predicción del abandono** - Desarrollado en detalle en APÉNDICE A (batch)
- **Predicción de lesiones** - (batch)

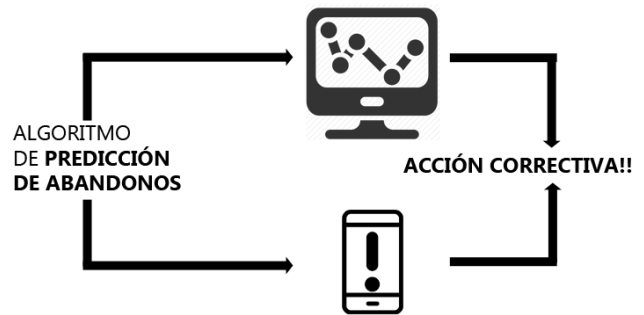


Predicción del abandono

El modelo de abandono tiene como objetivo el poder determinar la probabilidad de abandono, en los próximos meses, de cada uno de los usuarios del gimnasio, atendiendo a un conjunto de variables que describan a cada usuario y de datos históricos del gimnasio. Este modelo sería una herramienta de importantísimo valor para el gimnasio con la cual decidir qué estrategia tomar con cada usuario, de acuerdo con el riesgo de abandono y con el valor actual y potencial del usuario.

Al no disponer de datos reales de un gimnasio, se procede a construir una base de datos artificial que sea lo más realista posible, con el objetivo de obtener un modelo prototipo. Dicho modelo sería objeto de modificaciones una vez se tuviera una base de datos real, eliminando las hipótesis simplificadoras que permiten su entrenamiento y ajuste. Si bien no se dispone de información real suministrada por un gimnasio, sí que existe suficiente información publicada, tanto en artículos científicos como informes del estado del sector e informes publicados por consultoras especializadas en gimnasios low-cost, como para poder construir una base de datos próxima a la realidad.

La base de datos consta de C clientes que evolucionan en el tiempo, observándose su ingreso o alta y, en algunos casos, su abandono durante el periodo histórico que se utiliza para construir el modelo (36 meses o 12 periodos trimestrales). 6885 altas se observan en ese tiempo (este valor se justifica en la sección Modelo de abandono), que sumadas a los clientes ya



CLUSTERING DE CLIENTES + OBJETIVO
HISTORIAL DE ASISTENCIA/EJERCICIO
CARACTERIZACIÓN INTERES CLIENTE $C(i)$
- TIEMPO ESTIMADO PERTENENCIA EN GYM
- DETERMINACIÓN VARIABLES CARACTERÍSTICAS

Conceptualización del sistema de alerta por predicción de abandono

existentes en el periodo 1 (2700), hacen un total de $C = 9585$ clientes analizados en la base de datos.

Generalmente, el modelo de abandono va acompañado de un análisis descriptivo de la muestra utilizada para su construcción, dividiendo el total de la base de datos en dos o tres subconjuntos, para análisis, entrenamiento y validación. Dado que la base de datos es creada por nosotros, se obvia el análisis descriptivo, al quedar marcada por las hipótesis o condiciones que utilizamos para su construcción, y también la segmentación de los clientes. Al no disponer de una base de datos demasiado amplia, se decide que en vez de dividir la base de datos en dos conjuntos, se utilicen diferentes periodos para el entrenamiento y la validación, siguiendo la metodología de [6].

En el **APÉNDICE A** se describe cómo se construye la base de datos y qué variables caracterizan a cada uno de los usuarios finales del gimnasio, y cómo se obtiene el modelo de abandono, analizando al mismo.

[DATOS NO ESTRUCTURADOS]

Una de las grandes ventajas del enfoque y de la tecnología Big Data es que posibilita el análisis de datos no estructurados. Dentro del contexto de este proyecto, aprovecharemos tres grupos distintos de datos no estructurados. Aplicaremos técnicas innovadoras de data discovery y text mining para sacar el valor potencial que estos datos ofrecen. Es cierto que estos datos no son centrales para el modelo de negocio, pero también es cierto que ofrecen y permiten llegar a un entendimiento mayor del cliente y mejorar el servicio que se le ofrece. A continuación comentaremos los grupos de datos no estructurados susceptibles de análisis, y las preguntas más importantes que nos ayudarán a contestar.

Redes sociales

Las redes sociales hoy en día permiten a los usuarios comentar y compartir sus opiniones acerca de los productos y servicios que experimentan. Esta tendencia no para de crecer y cada vez más personas se unen al movimiento. En las redes sociales podemos encontrar valor para aplicar a nuestro negocio y ese valor se puede extraer basándose en técnicas como el Text Mining, desarrollando una aplicación concreta para cada caso de uso.

A través del análisis de las redes sociales podemos plantear y obtener respuesta a las siguientes preguntas:

- **Análisis de Sentimiento** - ¿Cómo perciben los usuarios nuestros productos?
- **Detección de puntos fuertes** - ¿Qué servicio valoran más nuestros usuarios?
- **Detección de puntos débiles** - ¿Dónde tenemos que mejorar?
- **Detección de oportunidades** - ¿Qué es lo que quieren nuestros usuarios?
- **Comparación con la competencia** - ¿Qué hacemos mejor? ¿Qué hacemos peor?
- **Posicionamiento en el mercado** - ¿Cómo estamos posicionados?

Beneficios:

Esta información permite conocer el negocio desde el punto de vista de los usuarios. Las respuestas a estas preguntas pueden ser vitales para asegurar el desarrollo y el crecimiento de Fitness Intelligence. Sirve para detectar tendencias y para detectar posibles fallos antes de que se produzcan resultados negativos en el negocio.

Datos meteorológicos

Este grupo de datos no estructurados son muy comunes a los proyectos de Big Data. En el contexto de Fitness Intelligence, dichos datos nos servirán a la hora de parametrizar y modelar el comportamiento de los usuarios.

El clima afecta en muchas maneras a la vida cotidiana de las personas. Desde el estado de ánimo hasta las saturaciones de las vías de transporte y la elección de como pasar las horas de ocio, son todas afectadas por el clima y el tiempo que hace.

En Fitness Intelligence queremos entender cómo afectan a nuestros usuarios el clima y los datos meteorológicos. Queremos saber si las variaciones en el tiempo afecta al entrenamiento de nuestros usuarios, el tiempo y frecuencia con la que asisten al gimnasio, su rendimiento en general y la motivación con la que afrontan el ejercicio

APIs que vamos a utilizar

Vamos a usar la API de twitter para saber qué opina la gente de los gimnasios que tienen implantado nuestro servicio y de los que aún no lo tienen. También usaremos una API que nos de información meteorológica (openweathermap.org) ya que esto repercute en la ocupación de los gimnasios, teniendo este dato nos será más sencillo optimizar las rutas de entrenamiento de los clientes.

Twitter: Utilizaremos esta API para conocer el feedback de los usuarios de nuestros clientes y prospects, tanto en tiempo real como de análisis convencional.

Estableceremos la comunicación mediante el lenguaje Python importando las librerías `sys`, `json` y `request`.



Twitter tiene 3 APIS diferentes (REST, Streaming y Ads), de las cuales utilizaremos principalmente:

- **API REST:** Mediante esta API obtendremos los datos referentes a los usuarios más relacionados con el fitness y la localización de los mismos para el análisis de sus quejas e ideas para la mejora del servicio ofrecido por los gimnasios a los que están apuntados. Con esta información, nuestro equipo comercial podrá argumentar de una manera mucho más precisa todas las ventajas que les podemos ofrecer, además nos aportará nuevas ideas para la mejora en términos generales del servicio que ofrecemos.
- **API Streaming:** Podremos saber en tiempo real si algún usuario de nuestros clientes está twitteando positiva o negativamente el servicio recibido. Si algo se puntúa negativamente y el usuario aún está en sala, el monitor podrá ayudarle para solucionar la incidencia.

Utilizaremos la versión free de la API, a continuación se puede observar un listado de peticiones Get que se nos permitirá hacer gratuitamente separadas por tipos:

<https://dev.twitter.com/rest/public/rate-limits>

Estas APIS nos devolverán un fichero JSON (clave, valor), que se almacenará en la base de datos de mongoDB, desde ahí con la API de SPARK y mediante las diferentes herramientas de nuestra arquitectura iremos dando forma a los datos.

Openweathermap

La versión free de esta API nos permite hacer hasta 60 peticiones Get por minuto, lo cual nos es más que suficiente para prever el incremento y la disminución de usuarios que habrá en sala por razones meteorológicas.

En esta ocasión también estableceremos la comunicación mediante Python, para ellos tendremos que importar la librería pyowm. Hay 9 APIS diferentes (<http://openweathermap.org/api>), pero en principio sólo utilizaremos dos:

- **Current weather data:** Nos permite hacer la llamada por el nombre de la ciudad, coordenadas geográficas, código postal, etc y nos responderá una lista de resultados de datos meteorológicos que coincidan con la búsqueda.

```
from pymongo import MongoClient

import twitter

# Datos de autenticacion

OAUTH_TOKEN='Rellenar con vuestros datos';
OAUTH_SECRET='Rellenar con vuestros datos';
CONSUMER_KEY='Rellenar con vuestros datos';
CONSUMER_SECRET='Rellenar con vuestros datos';

# Conexión a la API

api = twitter.Api(consumer_key=CONSUMER_KEY,
                  consumer_secret=CONSUMER_SECRET,
                  access_token_key=OAUTH_TOKEN,
                  access_token_secret=OAUTH_SECRET);
```

- **5 day / 3 hour forecast:** Nos permite saber el tiempo que hará de aquí a 5 días en adelante, las actualizaciones son cada 3 horas, los criterios de búsqueda son prácticamente iguales que en el caso anterior.

En este caso también se nos devolverá un fichero JSON y la manera de gestionarlo será igual que en el caso anterior.

```
import pyowm #librería que permite la conexión con la API

owm = pyowm.OWM('your-API-key') #Objeto que tendrá la conexión con la API
forecast = owm.daily_forecast("Madrid,es") #Aquí hacemos la petición a la API
tomorrow = pyowm.timeutils.tomorrow() #Aquí hacemos la petición a la librería
forecast.will_be_sunny_at(tomorrow)

observation = owm.weather_at_place(Madrid,es')
w = observation.get_weather()
print(w)

w.get_wind()
w.get_humidity()
w.get_temperature('centigrade')
observation_list = owm.weather_around_coords(-22.57, -43.12)
```


Dietas Deportivas

Muchos estudios sobre nutrición y deporte llegan a la misma conclusión: la nutrición y el deporte van de la mano. No se puede atender a uno mientras se descuida el otro. Es importante tener esto en cuenta al estar el negocio en el sector de Fitness, ya que si no se realiza una visión holística y global del entrenamiento, no se conseguirán los objetivos propuestos y el usuario estará descontento. Una aproximación acertada para combinar las rutinas de entrenamiento con la nutrición de los usuarios provocará resultados positivos, y una mayor confianza en la marca de Fitness Intelligence.

Recogiendo los datos de nutrición de nuestros usuarios (preferiblemente vía App, o bien un cuestionario inicial al apuntarse al gimnasio) y cruzándolos con los resultados y el desempeño de su entrenamiento, definimos un perfil de usuario.

Con estos datos podremos:

- Atender a los objetivos y necesidades del individuo



- Elaborar una dieta equilibrada acorde al gasto energético y físico
- Tener en cuenta las necesidades nutricionales específicas
- Elaborar una tabla de alimentos y ejercicios físicos. Podemos apoyarnos en Watson Chef

Esto reportará una serie de beneficios diferenciales a nuestros usuarios:

- **Dieta y entrenamiento efectivos:** No se malgasta el esfuerzo
- **Seguimiento del progreso:** Vista con perspectiva
- **Planteamiento integral:** Maximización de los resultados



BIG DATA

Fitness Intelligence no sería posible, o al menos sería mucho más complicado de llevar a cabo, sin contar con las tecnologías de almacenamiento y procesado de grandes volúmenes de información y de alta velocidad que brinda el Big Data.

A continuación realizaremos un recorrido a través de la infraestructura que posibilita que este proyecto sea posible, parándonos en cada uno de sus componentes más importantes.



Arquitectura Big Data sobre la que descansa Fitness Intelligence

Kafka

Es un sistema de almacenamiento distribuido, particionado y replicado. Es muy rápido en lecturas y escrituras, por tanto es muy útil para comunicar streams de información generados a gran velocidad, por tanto será la herramienta que recepcione los datos desde la Wifi de los gimnasios. Mediante esta herramienta nos aseguraremos de que le llegarán los datos a Spark en paquetes bien alineados y de que lleguen a tiempo.

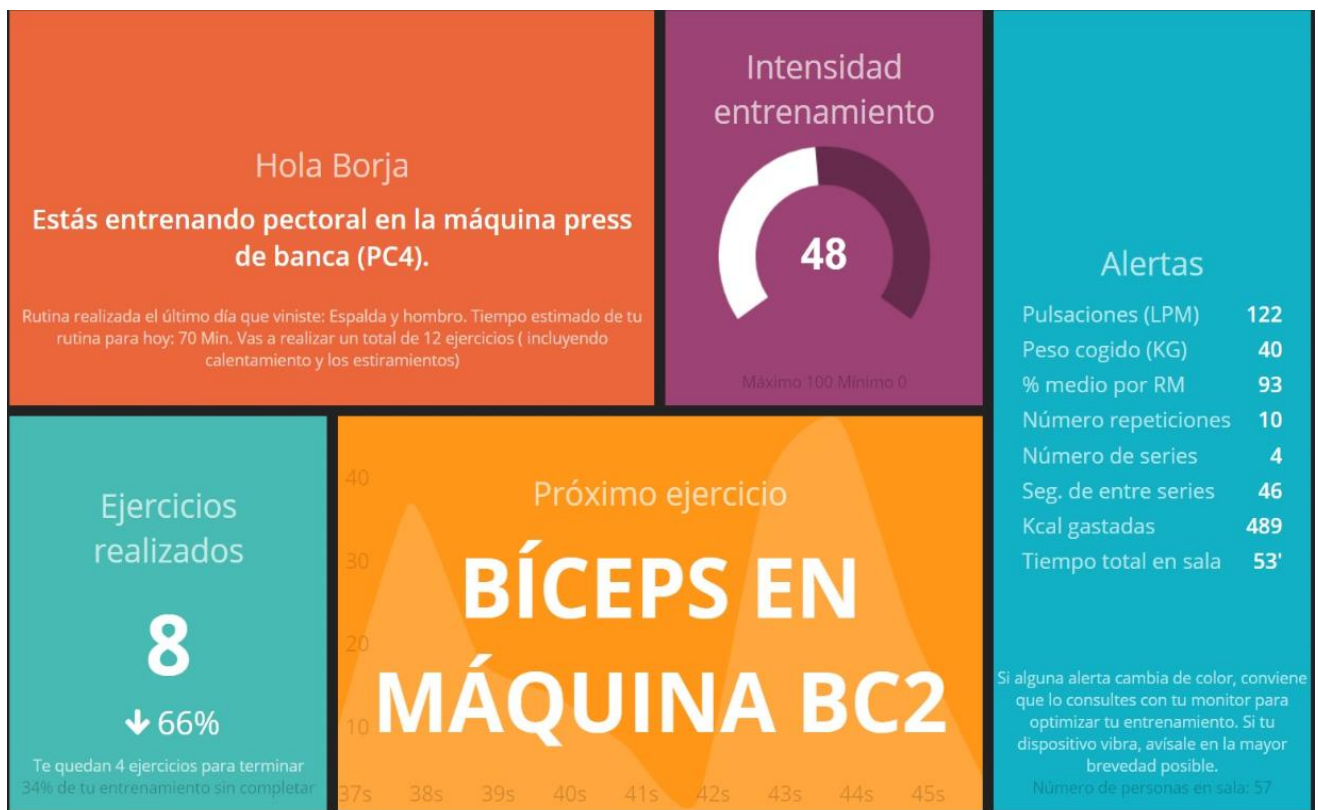
Spark Streaming

Spark ha sido diseñado para soportar en memoria algoritmos iterativos que se pueden desarrollar sin escribir en disco, un conjunto de resultados cada vez que se procese un dato. Va a ser la herramienta que utilizemos, ya que su velocidad nos permite trabajar en tiempo real (uno de los puntos fuertes

de nuestro proyecto). Además tiene un framework integrado para implementar análisis avanzados que incluye la librería MLib, la cual utilizaremos para los algoritmos machine learning que es una parte fundamental del servicio que ofrecemos. Una vez hayamos analizado nuestros datos mediante la librería MLib, los extraeremos mediante:

- **MongoDB:** Base de datos orientada a documentos. A diferencia de las bases de datos relacionales, con Mongo no tenemos que seguir un esquema, esto nos será muy útil debido a que los datos que solicitemos vía API nos llegarán en un JSON (clave-valor). MongoDB es una base de datos optimizada para gestionar este tipo de archivos.
- **Hive:** Ejecuta queries vía Apache Spark o MapReduce sobre datos almacenados en HDFS. Las consultas se ejecutan a través de HiveQL (las cuales se traducen automáticamente a MapReduce), es un lenguaje muy similar a SQL pero es más eficiente con bases de datos de gran tamaño.





Interfaz de usuario de Fitness Intelligence

Apache Cordova

Una vez gestionados los datos de los que acabamos de hablar, alimentarán parte de nuestra web y de nuestra app, la cual estará montada con apache cordova. Apache Cordova es un marco de desarrollo móvil de código abierto

Permite utilizar las tecnologías estándar web como HTML5, CSS3 y JavaScript para desarrollo multiplataforma, evitando el lenguaje de desarrollo nativo para cada plataforma móvil.

Utilizaremos esta herramienta ya que nos generará la App de forma automática partiendo de la página web de una manera compatible a todos los sistemas operativos, además lleva incluidas funcionalidades como las notificaciones Push, elemento fundamental para dar el servicio que ofrecemos a tiempo real.

Pentaho

Será la herramienta que utilizemos para

mostrar gráficas sobre el historial de los entrenamientos.

Nuestros clientes podrán actualizarla en cualquier momento y ver tanto gráficas en tiempo real del usuario, como de los usuarios de nuestros clientes. el histórico de los entrenamientos de los mismos.

Se alimentará tanto de datos extraídos mediante Hive como de la base de datos de mongoDB, para ello será necesario instalar Pentaho Report Designer.

Pentaho es un software libre bajo varias licencias, entre ellas la licencia GPLv2. Esta licencia implica que se puede usar, copiar, distribuir y modificar gratuitamente.

Esto es un punto a favor de este software ya que posibilita la implantación de un sistema BI bajo un desembolso nulo en licencias de software son caras debido a su complejo desarrollo.

Desde Hasta Hoy Desde Hasta Ahora



Usuarios en riesgo

Usuario	Puntuación
Borja	88
Dani	86
Cecilia	80

◀ Anterior Siguiente ▶



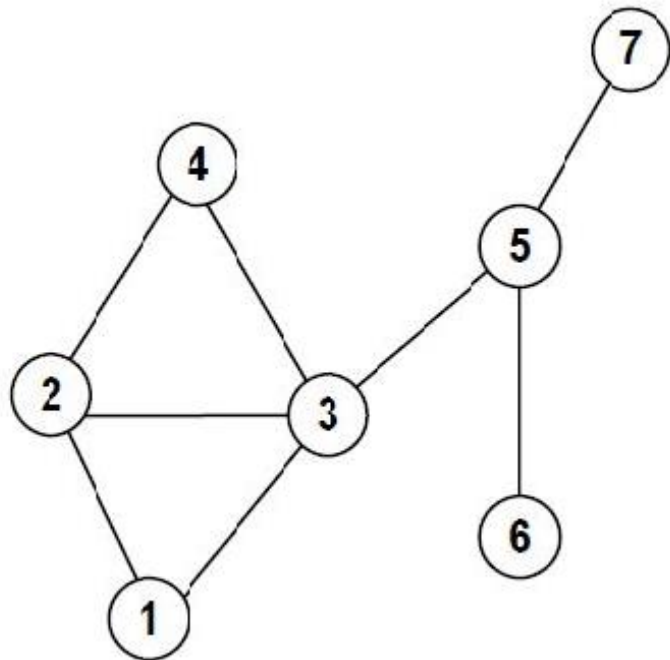
Panel de mando del entrenador

Desde nuestra página web, nuestros clientes metiendo su ID y su password (http://fitnessintelligence.eu/wp-login.php?redirect_to=http%3A%2F%2Ffitnessintelligence.eu%2Fwp-admin%2F&reauth=1), podrán acceder a diferentes gráficas de sus usuarios.

Neo4j

Es una base de datos orientadas a grafos (BDOG), ayuda a encontrar relaciones entre los datos y extraer su verdadero valor. Estas bases de datos tienen mejor rendimiento que las relacionales (SQL) y las no relacionales (NoSQL).

La clave es que, aunque las consultas de datos aumenten exponencialmente, el rendimiento de Neo4j no desciende. Gracias a esta herramienta, podremos hacer un grafo con la ruta a seguir los usuarios en los gimnasios.

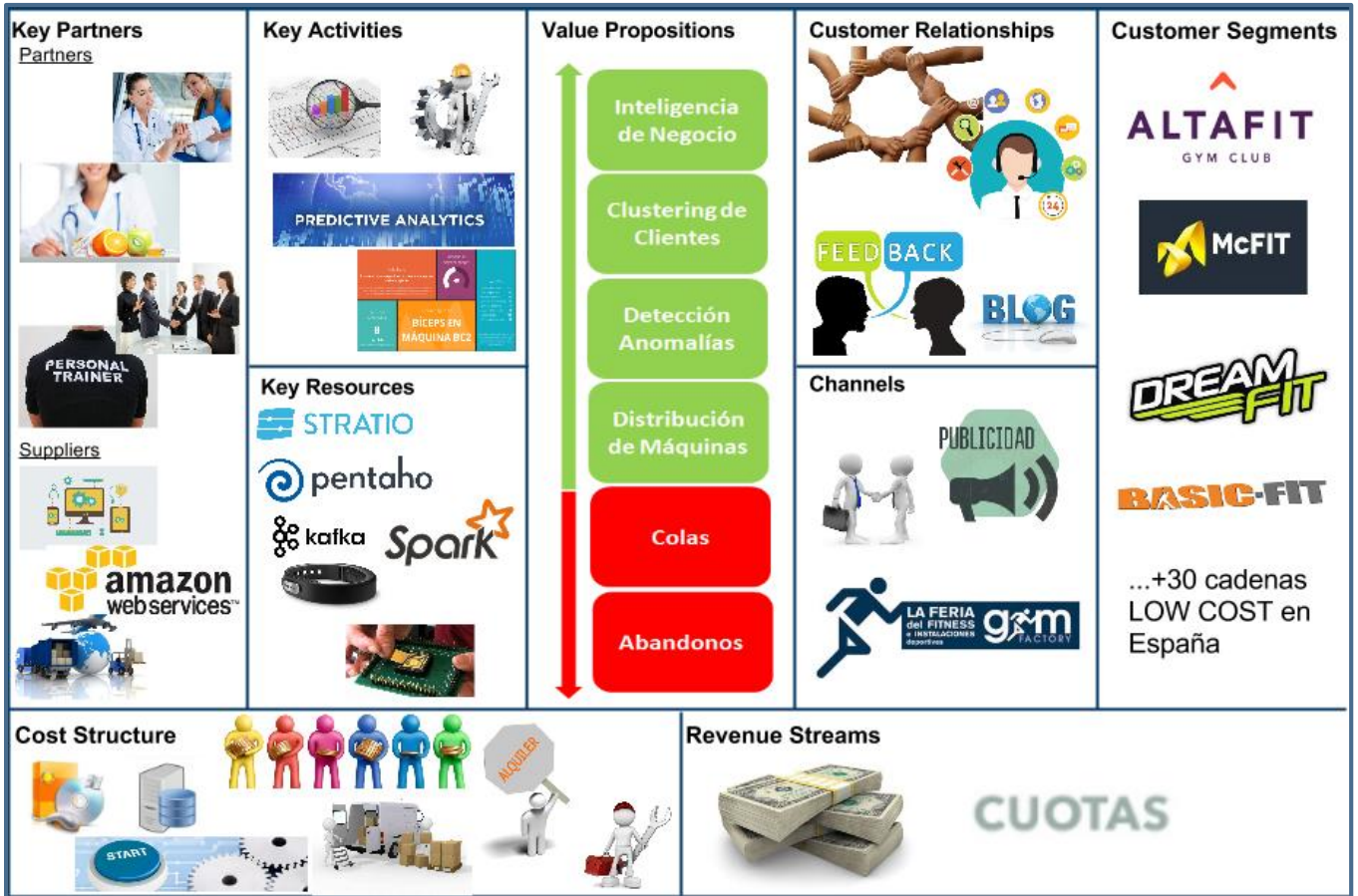


INVERSIONES / COSTES	PROYECTO GLOBAL			
	Fitness Intelligence			
	Año 1	Año 2	Año 3	Total
TOTAL INVERS. y COSTES	244.600 €	271.100 €	356.980 €	1.024.728 €
Gastos one-time	28.300 €	20.000 €	20.000 €	68.300 €
Crear sociedad limitada	3.300 €	- €	- €	3.300 €
Registrar patentes (Europa)	- €	20.000 €	20.000 €	40.000 €
Materiales para pruebas	10.000 €	- €	- €	10.000 €
Instalación y licencia Cloudera	15.000 €	- €	- €	15.000 €
Gastos recurrentes	216.300 €	251.100 €	336.980 €	956.428 €
Nóminas 6 empleados	108.000 €	118.800 €	130.680 €	501.228 €
Costes mantenimiento S.L.	2.300 €	2.300 €	2.300 €	9.200 €
Alquiler local IVIMA + suministros	3.000 €	4.000 €	5.000 €	18.000 €
Almacén (en el propio local)	- €	- €	- €	- €
Software ERP: Odoo (gratis)	- €	- €	- €	- €
Sensores y dispositivos de medición	- €	- €	- €	- €
Beacons y receptores RFID	6.000 €	28.000 €	100.000 €	134.000 €
Consumos de transporte, gasolina, etc.	1.000 €	2.000 €	3.000 €	6.000 €
Servicios Cloudera	66.000 €	66.000 €	66.000 €	198.000 €
Servidores Amazon	30.000 €	30.000 €	30.000 €	90.000 €

Business Case: Detalle de los gastos

INCREMENTO DE INGRESOS	PROYECTO GLOBAL			
	Fitness Intelligence			
	Año 1	Año 2	Año 3	Total
TOTAL D INGRESOS	40.000 €	280.000 €	1.000.000 €	1.320.000 €
Incremento de ventas	40.000 €	280.000 €	1.000.000 €	1.320.000 €
Gimnasios que contratan el 1er año (3) (ur	40.000 €	60.000 €	60.000 €	160.000 €
Gimnasios que contratan el 2º año (11)	- €	220.000 €	220.000 €	440.000 €
Gimnasios que contratan el 3er año (36) #	- €	- €	720.000 €	720.000 €

Business Case: Detalle de los ingresos



Business Model Canvas de Fitness Intelligence

Para cerrar este documento, aportamos el Business Model Canvas ilustrado del proyecto, que resume todo lo expuesto hasta ahora en el mismo.

Esperamos que hayan disfrutado del contenido. Si desean contactar con nosotros pueden hacerlo a través del email o de la web indicada en la primera página.

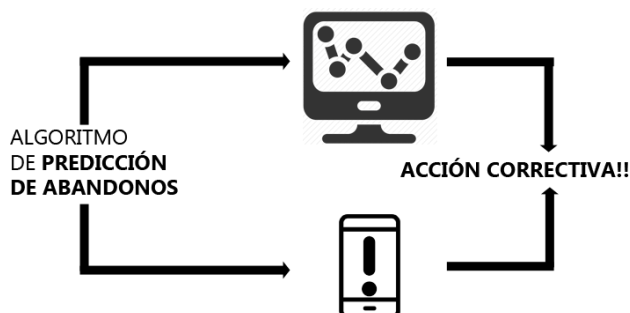


Predicción del abandono

El modelo de abandono tiene como objetivo el poder determinar la probabilidad de abandono, en los próximos meses, de cada uno de los usuarios del gimnasio, atendiendo a un conjunto de variables que describan a cada usuario y de datos históricos del gimnasio. Este modelo sería una herramienta de importantísimo valor para el gimnasio con la cual decidir qué estrategia tomar con cada usuario, de acuerdo con el riesgo de abandono y con el valor actual y potencial del usuario.

Al no disponer de datos reales de un gimnasio, se procede a construir una base de datos artificial que sea lo más realista posible, con el objetivo de obtener un modelo prototipo. Dicho modelo sería objeto de modificaciones una vez se tuviera una base de datos real, eliminando las hipótesis simplificadoras que permiten su entrenamiento y ajuste. Si bien no se dispone de información real suministrada por un gimnasio, sí que existe suficiente información publicada, tanto en artículos científicos como informes del estado del sector e informes publicados por consultoras especializadas en gimnasios low-cost, como para poder construir una base de datos próxima a la realidad.

La base de datos consta de C clientes que evolucionan en el tiempo, observándose su ingreso o alta y, en algunos casos, su abandono durante el periodo histórico que se utiliza para construir el modelo (36 meses o 12 periodos trimestrales). 6885 altas se observan en ese tiempo (este valor se justifica en la sección Modelo de aban-



CLUSTERING DE CLIENTES + OBJETIVO HISTORIAL DE ASISTENCIA/EJERCICIO
CARACTERIZACIÓN INTERES CLIENTE $C(i)$
 - TIEMPO ESTIMADO PERTENENCIA EN GYM
 - DETERMINACIÓN VARIABLES CARACTERÍSTICAS

dono), que sumadas a los clientes ya existentes en el periodo 1 (2700), hacen un total de $C = 9585$ clientes analizados en la base de datos.

Generalmente, el modelo de abandono va acompañado de un análisis descriptivo de la muestra utilizada para su construcción, dividiendo el total de la base de datos en dos o tres subconjuntos, para análisis, entrenamiento y validación. Dado que la base de datos es creada por nosotros, se obvia el análisis descriptivo, al quedar marcada por las hipótesis o condiciones que utilizamos para su construcción, y también la segmentación de los clientes. Al no disponer de una base de datos demasiado amplia, se decide que en vez de dividir la base de datos en dos conjuntos, se utilicen diferentes periodos para el entrenamiento y la validación, siguiendo la metodología de [6].

A continuación se describe cómo se construye la base de datos y qué variables caracterizan a cada uno de los usuarios finales del gimnasio, y cómo se obtiene el modelo de abandono, analizando al mismo.

Variable	Naturaleza	Valor
Sexo	Categórica	Hombre, Mujer
Edad (años)	Número continuo	18.0 – 75.0
Altura (cm)	Número continuo	157 – 200
Peso (kg)	Número continuo	50-130
Nivel educativo	Categórica	Primaria, Secundaria, Grado, Posgrado
Membresía (meses)	Categórica	<3, 3 - 6, 6 - 12, 12 – 24, >24
Frecuencia de asistencia mensual	Número entero	1-40
Duración de visitas (horas)	Número continuo	0.0 - 4.0
Veces de contacto con personal de gym	Número entero	0 – 4
Objetivo	Categórica	(ver tabla T4)
Experiencia anterior	Categórica	Mismo, Diferente, Ninguno
Tarifa (euros)	Categórica	24.90, 22.90, 19.90, 29.90, 26.90, 23.90

Tabla T. Descripción de las 15 variables que describen a un usuario de un gimnasio y que son utilizadas para nuestro modelo de abandono.

Base de Datos

La descripción de todos y cada uno de los usuarios finales de un gimnasio se realiza mediante 12 variables, las cuales se pueden agrupar en cuatro grandes bloques: variables socio-demográficas, variables de uso del gimnasio, expectativas y tarifa mensual. En la tabla T enumera cada una de las 12 variables, indicando su naturaleza y los valores que pueden tomar.

De acuerdo con [1], la distribución de usuarios finales de gimnasios low-cost por sexo es de un 53.5% de mujeres frente a un 46.5% de hombres. Este valor queda confirmado también por las referencias [3,4]. Si se analiza además la distribución por edades, los valores que se obtienen para cada segmento y por sexo son los recogidos en la tabla T1.

Intervalo edad (años)	<20	21-30	31-40	41-50	51-60	>60
% total usuarios	5.7	33.6	29.7	21.0	8.2	1.8
% total mujeres	6.2	36.2	27.7	20.9	7.8	1.2
% total hombres	5.2	30.6	32.1	21.2	8.4	2.5

Tabla T1. Distribución en % de los usuarios de un gimnasio low-cost por sexo y edad, [1].

Dado que del resto de variables no se dispone de más información que la relativa a los intervalos de edad y/o del sexo, cada una de las trece variables restantes se analiza como dependientes de edad y/o sexo e independientes de las restantes. Si bien es posible que exista una relación a priori se supondrá dicha simplificación.

La distribución de estatura, peso e índice de masa corporal (IMC) se ha construido basado en datos estadísticos publicados por instituciones públicas. Se considera que los valores de estatura y peso a incluir en la descripción del usuario del gimnasio corresponden al momento de alta del usuario. La estatura sí sigue una distribución normal si se analizan hombres y mujeres por separado [7], de modo que los porcentajes anteriores para cada sexo sirven para construir la distribución de estaturas donde, para los hombres variará como $N(178,10)$ y para las mujeres como $N(165,9)$. Mientras, el peso sigue una distribución lognormal de acuerdo con los valores de media y desviación típica sugeridos por [8,9].

En vez del IMC actual, se considera que resulta más significativo y útil manejar como variable el incremento de IMC, definido como la diferencia del IMC actual menos el IMC inicial. Mientras que el IMC inicial se calcula como el cociente del peso (realmente masa) por la estatura al cuadrado en el instante inicial o fecha de alta, el IMC actual se obtiene aleatoriamente a partir de una distribución binomial de acuerdo con los valores indicados en la tabla T2.

Otra variable importante sería el empleo de los usuarios finales, el cual puede condicionar su disponibilidad o actividad física diaria fuera del centro deportivo. Dado que dicha información no es disponible, y sí que puede estar correlada con la formación académica, al disponer datos de la distribución de estudios por

Cat. IMC	V	M
Bajo peso (16.0-18.5)	0.6	2.3
Normopeso (18.5-25.0)	55.9	72.0
Sobrep. I y II (25.0-30.0)	36.1	20.6
Obes. I y II (30.0 – 40.0)	7.4	5.1

T2. Frecuencia en % de categorías de IMC según la SEEDO por sexo, [3]

clientes de un gimnasio, se toma esta información para modelar el segmento de empleo en el cual se encuentra un usuario final. Cabe esperar que el nivel de educación pueda estar relacionado con un conjunto de empleos con exigencias semejantes. En [1] se indica la distribución de niveles de educación para usuarios de un gimnasio, siendo un 0.4% de los usuarios que tienen como nivel de educación más elevado la educación primaria, un 28.2% con bachillerato, un 56.9% con un grado universitario, y un 4.5% con un posgrado. Para modelar esta variable se utiliza una distribución multinomial.

El segundo bloque de variables hace referencia a las asociadas al uso o relación del usuario con el gimnasio. Este bloque incluye cinco variables: fecha de alta, membresía, frecuencia de asistencia semanal, duración de cada visita y contacto con personal del gimnasio en sus visitas. La fecha de alta y membresía, fundamentales para determinar si un usuario ha abandonado o no en los datos históricos del gimnasio, se analizan en la sección Modelo de abandono. En relación a la variable que representa la frecuencia de uso por semana, en [1] no se incluye información suficiente para construir la distribución correspondiente. Sin embargo, en [4] se presenta la frecuencia de uso por semana en un gimnasio privado. Igualmente, en [4] se incluye información

Edad (años)	0 veces	1 vez	2 veces	3 veces	> 3 veces
< 20	19.5	24.5	1.4	34.4	20.2
21 – 30	24.5	30.6	2.5	24.4	21.0
31 – 40	22.4	30.3	2.2	22.2	22.9
41 – 50	19.8	29.7	2.1	23.1	25.3
51 – 60	18.5	29.6	1.9	26.2	23.8
> 60	18.3	24.4	2.8	30.5	23.0

T3. Contacto con personal del gimnasio semanalmente

sobre la duración en horas de cada visita al centro. Al ser información general (% del total de usuarios que acuden con una frecuencia f o cuya visita dura un tiempo t), se elige una distribución normal para ambas variables. Efectivamente, si se muestran los histogramas de estas dos variables, se observa cómo se ajustan a una distribución normal $N(3.7, 2.1)$ y $N(1.5, 0.6)$, respectivamente.

Otra variable que puede afectar la retención o abandono de un usuario final en un gimnasio es su contacto con personal del gimnasio. Así, en [1] se incluye una distribución por edades para diferentes grados de contacto con personal del gimnasio para cada vez que asistía al centro. Esta información se muestra en la tabla T3.

Estos datos de porcentajes son usados para crear los valores de la variable contacto de la base de datos, siguiendo para cada segmento de edad una distribución multinomial, donde cada intervalo de permanencia tiene una probabilidad de acuerdo a los porcentajes mostrados en la tabla T3.

Finalmente, las variables dentro del bloque expectativas son el objetivo inicial y la experiencia anterior. Para el objetivo inicial, se utiliza la información publicada en [4], donde se recoge, en porcentaje, qué objetivo o motivación tenían los usuarios de un gimnasio para apuntarse a ese centro, datos que se muestran aquí en la tabla T4

Cada una de las motivaciones, independientemente de qué combinación de objetivos individuales lleve asociada, se considera como una categoría diferente. De este modo, la variable objetivo inicial queda representada mediante una distribución multinomial, donde cada probabilidad viene dada por el porcentaje correspondiente de la tabla T4.

Objetivo	% poblac.
Fitness	44.7
Salud	21.0
Fit. – cont. peso –relax	9.2
Control peso	6.6
Relax	5.3
Salud – fitness	5.3
Otro	2.7
Fitness – control peso	2.6
Fitness – relax	2.6
Socializar	0.0

T4. Objetivos

Para la experiencia anterior, se utiliza la información publicada en [1], siendo la distribución por edades de usuarios que ya habían estado en el mismo centro, que habían estado en un centro diferente o que no habían asistido a ningún centro antes de acudir al actual los mostrados en la tabla T5.

Intervalo edad (años)	<20	21-30	31-40	41-50	51-60	>60
Mismo centro	6.1	37.7	28.8	17.3	8.3	1.8
Centro diferente	4.0	33.4	31.5	21.4	8.0	1.8
Ninguno	11.0	33.4	24.3	20.4	8.3	2.5

Tabla T5. Experiencia previa a la actual.

Se sabe que el 72% de los usuarios había estado inscrito en otro centro de fitness antes del actual, mientras que un 4.6% había estado en el mismo centro, y un 23.4% no había estado nunca en un centro de fitness. Con estos datos, la tabla T5 se puede reescribir para mostrar la distribución de experiencia segmentada por edades, igual que en los casos anteriores. Los porcentajes resultantes son usados, a su vez, para crear los valores de la variable experiencia anterior de la base de datos, siguiendo para cada segmento de edad una distribución multinomial.

Finalmente, a cada usuario del gimnasio se le asigna una tarifa de acuerdo con su vinculación y con el objetivo que se marca en el momento de alta. A fin de simplificar el estudio, se consideran seis tarifas diferentes, correspondientes a un abono mensual, trimestral o anual, y a si el abono es básico (fitness, relax, control de peso, salud) o si contempla clases colectivas (en este caso se traduce en una combinación de las anteriores). Los valores de las tarifas que se consideran aquí varían entre 19.90 y 26.90 €/mes.



Modelo de abandono

Para el modelo de abandonos, se considera como período histórico 36 meses desde la fecha actual (julio 2014 - mayo 2016), dividiendo este tiempo en 12 períodos, de tres meses de duración cada uno. En cada uno de esos 12 períodos, se observan clientes nuevos y clientes que abandonan el gimnasio. A pesar de la evidente influencia del momento del año (inicio de año o verano), como hipótesis simplificadora consideraremos que el número de altas se mantiene uniforme en el tiempo, mientras que el abandono se produce de acuerdo a la información publicada en [1,2] sobre cuántos meses llevan los usuarios de un gimnasio como socios en la fecha actual. Esta última información necesita ser procesada para ver cómo traducir el tiempo de permanencia como socio en tasa de abandono mensual o por trimestre.

Así, se considera una distribución uniforme con probabilidad $1/12$ para las altas en el gimnasio, habiendo N_{in} nuevos clientes en cada periodo y siendo el total de altas en un año un 85% del número medio de usuarios del gimnasio, de acuerdo con el dato de incremento anual de los ingresos en un gimnasio low-cost, [3] corregido por el número de abandonos de clientes nuevos que se observa en un año (aproximadamente 75%). Las altas se suman a los N_{gym} usuarios ya existentes en

Edad (años)	<3 meses	3-6 meses	6-12 meses	12-24 meses	>24 meses
< 20	37.2	22.3	16.1	19.1	5.3
21 – 30	23.7	19.0	20.2	24.5	12.6
31 – 40	17.2	16.0	19.8	27.3	19.7
41 – 50	16.2	12.7	18.6	27.9	24.6
51 – 60	16.8	13.0	16.5	26.5	27.2
> 60	12.2	8.3	20.6	29.4	29.4

T3. Contacto con personal del gimnasio semanalmente

el gimnasio, y dan como resultado los $N(t)$ clientes del gimnasio en cada periodo t . A fin de dotar de mayor realismo al número de usuarios dados de alta en el gimnasio en un periodo, se podría considerar una distribución normal con media 2650 y desviación típica 318, correspondiente al 12% de los usuarios que abandonarían el gimnasio en un mes. La diferencia entre dos períodos consecutivos dan como resultado el número de bajas en el período previo, sin contabilizar las altas del periodo final.

Las referencias [1,2] analizan la permanencia de los usuarios de un gimnasio en el centro. La información publicada sobre qué porcentaje de los usuarios llevan un tiempo T en el centro es usada para construir un histograma de permanencia en el gimnasio. En la tabla T6, se indica el porcentaje de los usuarios correspondientes a un segmento de edad que han permanecido un tiempo T , donde ese tiempo T se discretiza en varios intervalos.

De este modo, es posible para cada segmento de edad representar el histograma de permanencia en el gimnasio. Teniendo en cuenta que los períodos indicados no son uniformes, dado que unos corresponden a 3 meses, y otros a 6 ó 12 meses, se supone una distribución uniforme dentro de cada intervalo. Consecuentemente, si se representa la curva acumulada de usuarios que al menos llevan un tiempo T , se puede ver cómo, por

ejemplo, si se toman datos mensualmente, el 62.8% de los usuarios actuales llevan al menos 4 meses.

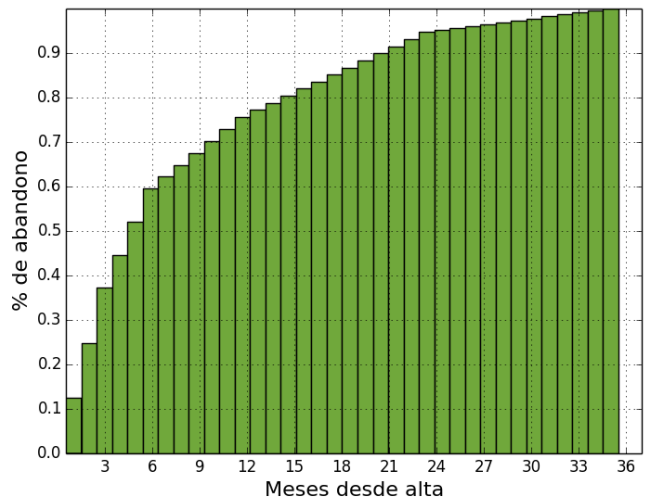


Fig. F1. Curva de probabilidades acumulada de abandono en un gimnasio

Para continuar con la construcción de la base de datos, si se vuelve a apelar a la hipótesis de que no existen variaciones en el número de altas a lo largo del año, la evolución de permanencia nos permite obtener la evolución de abandono. La figura F1 muestra la curva de probabilidades acumulada a partir de los datos de la tabla T6, observándose cómo aproximadamente el 50% de los usuarios habrían abandonado el gimnasio en los primeros 6 meses, dato que es confirmado por la referencia [3]. Con esta información, es posible determinar, para cada periodo, cuántos miembros del gimnasio se dan de baja, pudiendo además estimar el porcentaje de usuarios por edad y sexo. Es decir, para cada período se tiene que cada usuario no sólo queda definido

por las variables indicadas en la tabla T, sino además una variable de indicación de abandono, tomando el valor 1 si sí ha abandonado y 0 en el contrario. Los valores de este indicador de abandono se asignan en nuestra base de datos artificial de acuerdo con el valor $N(t+1) - N(t) - N_{in}(t+1)$ y las probabilidades indicadas en la figura F1 que se utilizan para una distribución multinomial.

Como se ha mencionado anteriormente, se construye una base de datos de 12 periodos durante los 36 meses anteriores a la fecha actual. Sin embargo, si se utilizaran todos los periodos, de acuerdo con el número de clientes medio por periodo (2650), tendríamos una base de datos muy grande para sólo entrenar el modelo. Dado que no sólo es necesario su entrenamiento, sino su validación (no se considera ajuste al trabajar con un modelo de regresión logística), se toman 8 de los 12 períodos como base de datos para entrenar el modelo, y los 4 restantes como validación. Los 8 períodos que se eligen para el entrenamiento del modelo son los que van de 3-10 (i.e. diciembre 2014-febrero 2016).

Se ha considerado un modelo de regresión logística con una variable dependiente binaria. El modelo se estima utilizando todas las variables de la tabla T, pero aquellas variables que no son significativas son eliminadas en el modelo. El proceso de selección de las variables se realizó de manera iterativa, donde en cada iteración se analizaba el comportamiento estadístico de cada variable, y sólo aquellas variables que tuvieran un valor p inferior a 0.05 se consideraban significativas. Una vez reducido el número de variables, el modelo se volvía a ajustar y se repetía el análisis de significación estadística de cada variable, repitiendo este proceso hasta que todas las variables que utilizamos en el modelo tienen un valor p inferior a 0.05. El resultado final de las variables utilizadas en el modelo aquí presentado es el mostrado en la tabla T7.

La mayor parte de las variables del bloque sociodemográfico han desaparecido, quedando sólo la edad, observándose que la importancia del sexo se diluye después de varias iteraciones. En el bloque de variables asociadas a la relación del usuario con el gimnasio, se observa que sólo la duración de la visita desaparece como variable del modelo, pudiéndose explicar por la poca variabilidad de las opciones y que el 74.0% de los usuarios acuden entre 1 y 2 horas. En el tercer bloque de expectativas, sorprendentemente desaparece la variable objetivo. La tarifa también ha desaparecido, siendo esto algo que se contradice con lo publicado por [3], donde muy probablemente debido a la poca diversidad de tarifas contempladas en nuestra base de datos y a que el valor de la variable tarifa estaba definida como función de la membresía y, por lo tanto, se entiende que quedaba correlada con ella.

Variable	Valor
Edad (años)	0.0017
Membresía (meses)	0.0002
Frecuencia de asistencia mensual	0.0138
Veces de contacto con personal de gym	0.0053
Experiencia anterior	0.0190

T7. Variables consideradas para el modelo de regresión logística.

Cabe indicar que de las variables que han quedado, mención especial merece la variable experiencia anterior, al tratarse de una variable categórica. Para poderla procesar correctamente en la construcción del modelo de regresión logística, la variable se transforma en una variable dummy, la cual toma valores 0, 1 y 2 para cada uno de los niveles de la variable (otro

gimnasio, mismo gimnasio o ninguno). El valor de los coeficientes del modelo, para cada una de las variables y el término independiente, se muestra en la tabla Tb.

Objetivo	% poblac.
B ₀ (Término independiente)	0.363
B ₁ (Edad)	0.308
B ₂ (Membresía)	0.439
B ₃ (Frecuencia de asistencia semanal)	-1.323
B ₄ (Veces contacto con personal de gym)	-0.352
B ₅ (Experiencia anterior)	-0.017
B ₀ (Término independiente)	0.363
B ₁ (Edad)	0.308
B ₂ (Membresía)	0.439
B ₃ (Frecuencia de asistencia semanal)	-1.323

Tabla Tb. Coeficientes del modelo de regresión logística.

Así, el modelo resultante es

$$\ln[p/(1-p)] = b_0 + b_1 * Edad + b_2 * Membresía + b_3 * Frecuencia + b_4 * vecescontacto + b_5 * experiencia$$

Los resultados de las probabilidades continuas que se producen con el modelo de regresión logística se discriminan en dos grupos a partir de un valor de corte, siendo ese valor, por lógica 0.5, de modo que probabilidades superiores a 0.5 indican socios que abandonan el gimnasio, y valores inferiores o iguales a 0.5 indican que no abandonan. En la figura F2 se

muestra la predicción del modelo de regresión logística para el conjunto de entrenamiento. El valor dado por el modelo de regresión se representa para cada uno de los usuarios incluidos en el conjunto de entrenamiento (6209). Los símbolos representan el abandono real o no del usuario. Se observa claramente que la mayoría de los usuarios que abandonan realmente se encuentran en el segmento de probabilidades superior a 0.5, aunque el número de usuarios que abandonan y que el modelo predice una probabilidad inferior a 0.5 es todavía notable. La misma observación se obtiene para los usuarios que no abandonan.

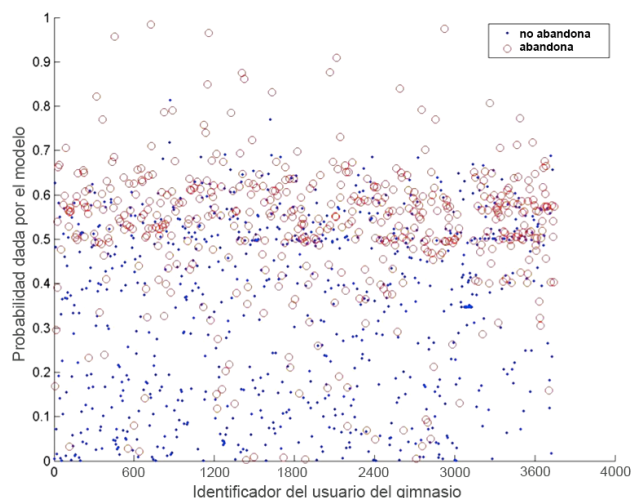


Fig.2. Comparativa de la predicción del modelo de abandono con el conjunto de puntos de entrenamiento.

Analizando la capacidad predictiva del modelo, si se mantiene como valor de corte el mismo que en el entrenamiento (0.5), la matriz de confusión obtenida es la mostrada en la tabla TN. El porcentaje de predicciones correctas del modelo es del 71%, sumando verdaderos positivos y verdaderos negativos. El % de usuarios que fueron identificados por el modelo como usuarios que abandonan y que realmente abandonaron es del 58.2% (verdaderos positivos), mientras que el % de usuarios que el modelo no fue capaz de predecir como usuarios que abandonan y que finalmente abandonaron es del 8.5% (falsos

		Base de datos. Conjunto validación		
		Abandonan	No abandonan	
Predicción del modelo	Abandonan	963	691	1654
	No abandonan	288	1434	1722
		1251	2125	3376

Tabla TN. Matriz de confusión

negativos). Más útiles son los parámetros de cobertura y tasa de error, que para este valor de corte toman los valores de 0.77 y 0.33, respectivamente. Como se podrá observar posteriormente en la figura F3, el modelo de regresión tiene una calidad media-alta en su capacidad de predicción, a pesar de que el valor de la eficiencia (F-score), es del 66.1%.

La curva ROC, donde se muestra la cobertura o tasa de positivos verdaderos frente a la tasa de error o tasa de falsos positivos, para cada valor de corte, se muestra en la figura F3. La curva del modelo definitivo se representa en negro, incluyendo las curvas de los modelos obtenidos en iteraciones previas hasta reducir el número de variables hasta aquellas con un valor p inferior a 0.05.

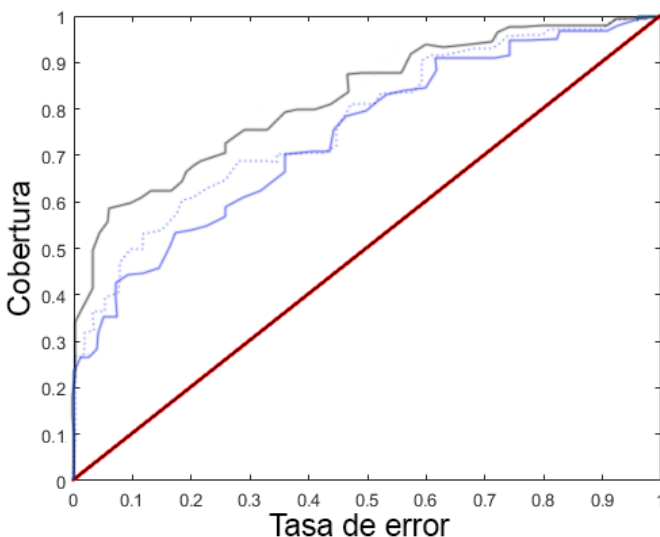


Fig. F3. Curva ROC del modelo de regresión logística final (negro), y de los modelos de regresión previos (azul continuo y azul punteado)

Bibliografía

- [1] Valcarce, M., López, F., García-Fernández, J., 4º Informe Gimnasios Low-cost en España, Valgo Fitness & Sport Management, enero 2016
- [2] García-Fernández, J., Fernández-Gavira, J., Bernal, A., La percepción de calidad y fidelidad en clientes de centros de fitness low cost, Suma Psicológica 21(2), pp. 123–130, 2014
- [3] Estudio anual del mercado del fitness en España en 2015. Contexto, situación y tendencias del sector en España, Life Fitness Zoom Mercado, enero 2016
- [4] Afthinos, Y., Theodorakis, N. D., Nassis, P., Customers' expectations of service in Greek fitness centers, Managing Service Quality 15(3), pp. 245–258, 2005
- [6] Ascarza, E., Hardie, B. G., Modeling churn and usage behavior in contractual settings, London Business School, 2009
- [7] A'Hearn, B., Peracchi, F., Vecchi, G., Height and the normal distribution. Evidence from Italian Military Data, Demography 46(1), pp. 1–25, 2009
- [8] Jolicoeur, P., Introduction to Biometry, Springer Science+Business Media, 1999
- [9] Burnmaster, D. E., Crouch, E. A., Lognormal distributions for body weights as a function of age for males and females in the United States, 1976-1980, Risk Analysis 17(4), pp. 499–505, 1997

