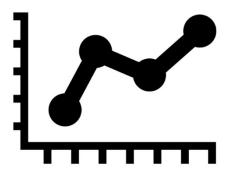
Caso 3: Activación de clientes dormidos

Descripción

El principal objetivo del presente caso de uso es calcular un índice de estimación de la recuperación de clientes dormidos.

El output del proyecto es poder diseñar acciones comerciales para recuperar clientes dormidos, generando un listado de clientes con mayor probabilidad de recuperación y por último proporcionar al cliente las herramientas de visualización dinámicas de resultados que les permitan identificar y actuar sobre los clientes dormidos de forma efectiva.





Caso 3: Activación de clientes dormidos

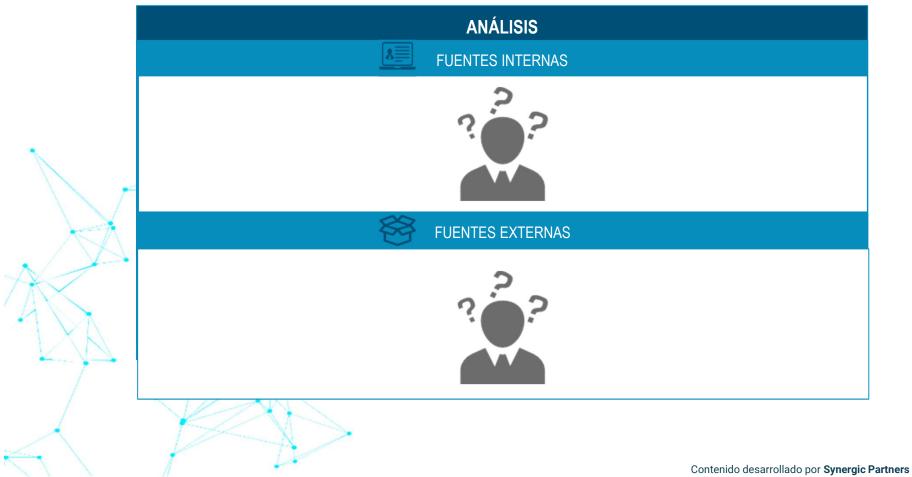
Valor de Negocio



Caso 3: Activación de clientes dormidos

Fuentes de datos

¿Qué fuentes de datos utilizarían para la realización del caso de uso?



Caso 3: Activación de clientes dormidos

Fuentes de datos

ANÁLISIS

FUENTES INTERNAS



- Data Warehouse: El data warehouse de la compañía posee la información de los recibos y las tarjetas de los clientes, tipología de operaciones en base al sector del comercio, esta información ayuda a determinar el gasto mensual en los diferentes sectores.
- **Datamart Comercial:** información estática de las diferentes entidades (personas, productos, oficinas, etc.). Además contiene los datos demográficos de los clientes, tablas contractuales, histórico de productos de cliente y tabla de comisiones entre otras.

FUENTES EXTERNAS



• Redes sociales: A través del email de los clientes se buscó su perfil en redes sociales. Sólo se tenía la dirección de correo del 26% de los clientes, y de estos sólo un 11% estaban presentes en las redes sociales, por lo que esta fuente de datos no se incluyó en el modelo. El objetivo era digitalizar las campañas para activación de clientes dormidos



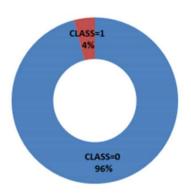
Caso 3: Activación de clientes dormidos

Tratamiento de datos

Balanceado de muestras

El conjunto de datos se encontraba desbalanceado, esto es, un alto porcentaje de datos de una clase frente a un pequeño porcentaje de datos de la clase contraria. Para una correcta ejecución del modelo primeramente hay que balancear los datos para que el porcentaje de ambas clases sea similar.

Clientes dormidos



Analítica	Algoritmo	Funcionamiento
Balanceado	RUSBoost	Algoritmo híbrido de sampling/boosting (RUS: Random Undersampling) Este algoritmo realiza un <i>under sampling</i> de la clase mayoritaria, en este caso la clase 0. Es decir, disminuye la cantidad de la misma para equipararla a la cantidad de datos de la clase 1. Permite el uso de diferentes modelos de clasificación.

Caso 3: Activación de clientes dormidos

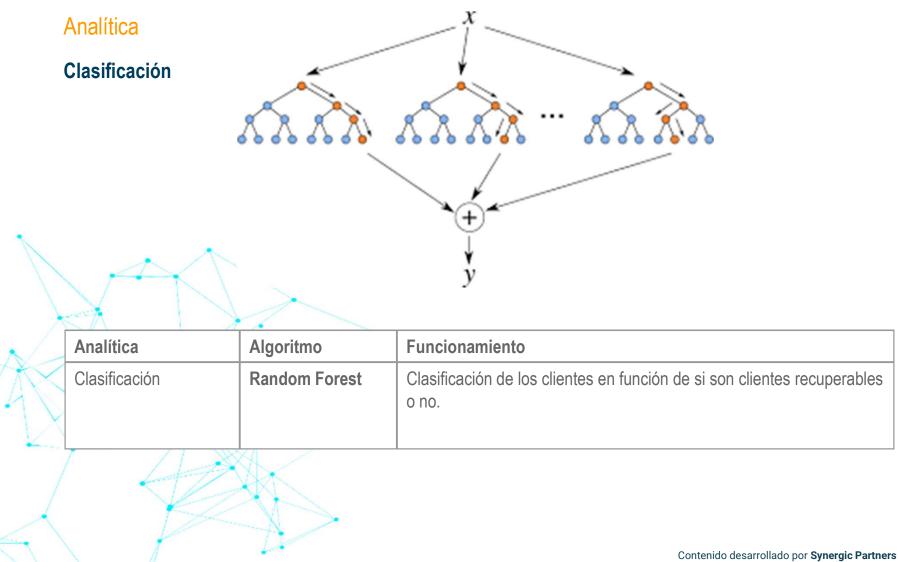
Analítica

¿Cómo lo resolvemos?

- Al no haber una gran cantidad de datos, la analítica fue local haciendo uso del lenguaje de programación R.
- Debido a un gran desbalanceo de los datos, se utilizó un modelo con under sampling.
- Además se realizó un cruce con datos de redes sociales con el objetivo de digitalizar las campañas de marketing.



Caso 3: Activación de clientes dormidos



Caso 3: Activación de clientes dormidos

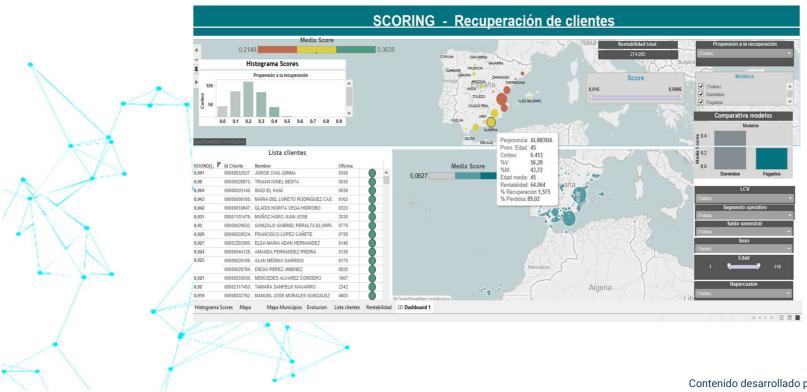
Resumen del proceso



Caso 3: Activación de clientes dormidos

Output

Gracias al diseño de un dashboard interactivo el cliente puede identificar fácilmente los clientes con mayor propensión a ser recuperados. El dashboard también aporta información relativa a la rentabilidad total esperada, datos cualitativos sobre el cliente y la zona geográfica donde se encuentra la oficina a la que pertenece el cliente.

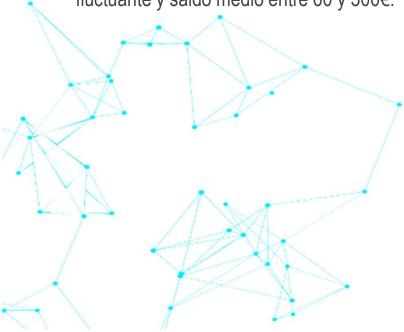


Caso 3: Activación de clientes dormidos

Resultados

Detectar los clientes dormidos permite realizar campañas comerciales personalizadas, **evitando** así la **fuga** de los mismos.

Estos clientes detectados como dormidos fueron principalmente **menores y jóvenes** , con saldo fluctuante y saldo medio entre 60 y 300€.







Caso 4: Propensión de contratación de cuenta nómina



El **objetivo** de este caso de uso es determinar la **propensión a la compra de la cuenta nómina** de los clientes de una entidad bancaria.



Este caso de uso es solicitado por el departamento de **CRM**, departamento de estrategia de negocios centrada en el cliente.



El departamento de CRM tiene un **área comercial**, la cual es **limitada**. No pueden abarcar contactar con todos los clientes para ofrecerles una cuenta nómina, por lo que necesitan **acotar** aquellos clientes propensos a contratarla.



- → Valor empresa: Focalizar los esfuerzos del departamento de marketing para obtener un mayor porcentaje de éxito.
- → Valor Cliente: El cliente evitará que se le incomode con campañas telefónicas en las que no muestre interés.





USE CASE:
PROPENSIÓN A LA CUENTA NÓMINA

Caso 4: Propensión de contratación de cuenta nómina

Descripción

El **objetivo** de este caso de uso es determinar la **propensión a la compra de la cuenta nómina** de los clientes de una entidad bancaria.

El departamento de CRM tiene un **área analítica** y un **área comercial**, la cual es **limitada**. No pueden abarcar contactar con todos los clientes para ofrecerles una cuenta nómina, por lo que necesitan **acotar** aquellos clientes propensos a contratarla. El área analítica no consigue dar con estos clientes de modo que se hace uso de tecnologías big data para solucionar el problema.

La idea consiste en generar un modelo analítico que dé como resultado una **lista de clientes** ordenados por **scoring** según la propensión.



Caso 4: Propensión de contratación de cuenta nómina

Valor de Negocio



Optimización de servicios

- Focalizar la función comercial.
- Evitar ejercer campañas de marketing y de contacto comercial telefónico a clientes sin interés en contratar el servicio, evitando la pérdida de tiempo y de recursos.



Revenue

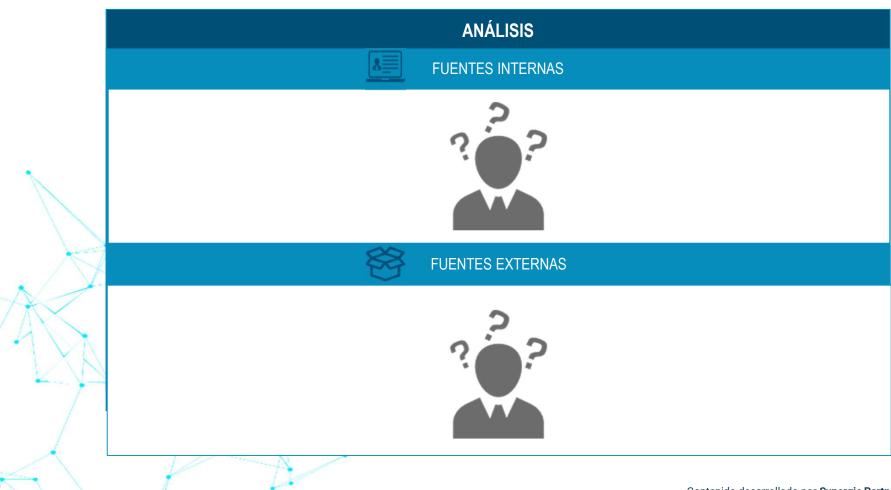
 Incrementar el revenue assurance mediante la identificación de clientes propensos a contratar el servicio, trayendo su cuenta nómina al banco,



Caso 4: Propensión de contratación de cuenta nómina

Fuentes de datos

¿Qué fuentes de datos utilizarían para la realización del caso de uso?



Caso 4: Propensión de contratación de cuenta nómina

Fuentes de datos

ANÁLISIS



FUENTES INTERNAS

- Datos internos clientes: Conjunto de 1900 variables agrupadas en 56 categorías: banca telefónica, cheques, créditos, variables sociodemográficas, depósitos, disposiciones, broker, financiación, fondos, grandes compras, incidencias, acciones comerciales, nómina, posiciones, préstamos hipotecarios, seguros, tarjetas de crédito, tarjetas de débito, transferencias, valores, etc
- Transcripciones conversaciones comercial-cliente.

器

FUENTES EXTERNAS

- Open Street Map: Datos geográficos para relacionar el código postal del cliente y las oficinas bancarias en esa zona.
- INE (Instituto Nacional de Estadística): Extracción por código postal de residencias, pisos, precios, actividad, comercios...
- Redes sociales: Búsqueda de publicidad de la entidad bancaria y la competencia donde se hace alguna mención a una cuenta nómina.



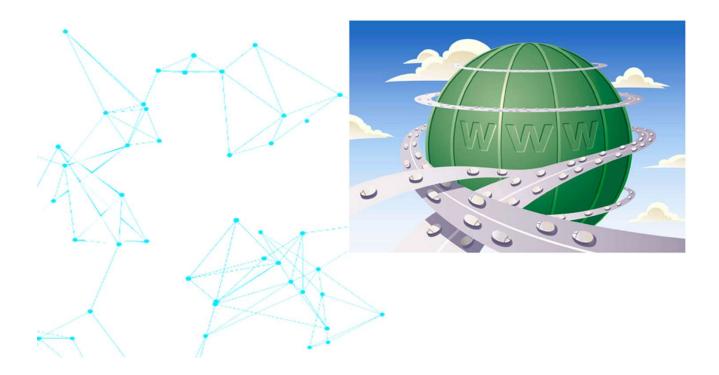
Caso 4: Propensión de contratación de cuenta nómina

Fuentes de datos

¿Otras fuentes?

Datos de **logs de navegación** dentro de la página interna del banco.

Datos sobre el tiempo que el cliente permanece navegando en la página, cantidad de veces que la visitan, por qué zonas de la página navegan, si acceden a consultar información sobre la cuenta nómina, etc.



Caso 4: Propensión de contratación de cuenta nómina

Tratamiento de datos

El conjunto de variables era muy extenso, 1900 agrupadas en 56 categorías.

De estas variables existentes se crearon 109 derivadas, que podemos clasificarlas en tres clases:

- Agregaciones: Agrupación de datos, por ejemplo, agrupar todas las acciones comerciales bajo una sola variable.
- Tendencias: Tendencia de los valores de una variable
- Ratios: Por ejemplo, porcentaje de las acciones comerciales mediante SMS respecto al total de acciones comerciales.

Además se incluyeron 19 variables de **fuentes externas** y text mining (transcripciones telefónicas). Del conjunto total se **seleccionaron** manualmente **373 variables**.



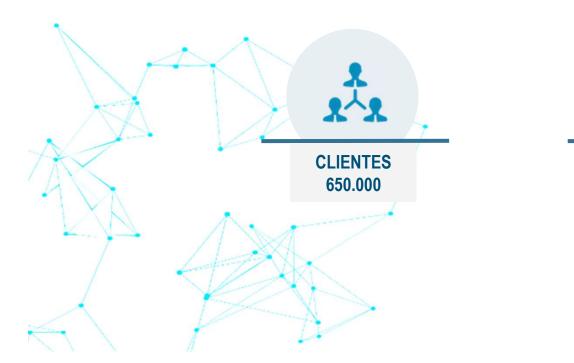
Caso 4: Propensión de contratación de cuenta nómina

Tratamiento de datos

La población objetivo tenía como características:

- No haber tenido cuenta nómina en los últimos dos años
- Podían tener o no ingresos regulares
- Se diferenció entre **cliente** (agrupación familiar, parejas...) y **personas**

Se seleccionaron tres meses de ventana para el modelo y los dos meses siguientes para la evaluación, es decir, **ventanas de 5 meses**.





Caso 4: Propensión de contratación de cuenta nómina

Analítica

¿Cómo lo resolvemos?

El proceso analítico para identificar aquellos clientes favorables a contratar una cuenta nómina fue el siguiente:



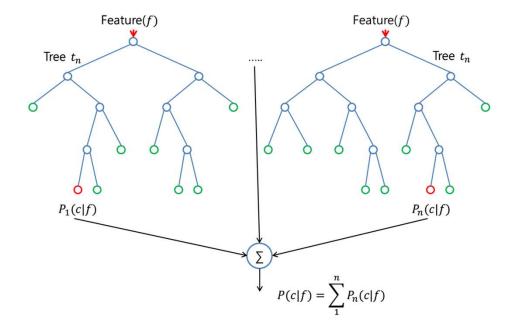
Caso 4: Propensión de contratación de cuenta nómina

Analítica

Clasificación

Se estudiaron **tres posibles modelos** de **clasificación**: Decision Trees, Random Forest y Logistic Regression.

Se hicieron pruebas con los tres para obtener el que daba mejores resultados, que resultó ser **Random Forest.** Con este algoritmo se buscaba clasificar si el cliente era propenso SI o NO a contratar la cuenta nómina.





Caso 4: Propensión de contratación de cuenta nómina

Analítica

Clustering

La gran cantidad de variables daba lugar a demasiados clusters, por lo que se llevó a cabo una segmentación dirigida de clientes en tres grupos:

- Clientes a **consolidar**: con ingresos regulares
- **Cancelados**: Clientes que habían tenido cuenta nómina y la habían cancelado en un tiempo anterior a dos años.
- Otros



Caso 4: Propensión de contratación de cuenta nómina

Analítica

Text Mining

Cuando un comercial ejerce una acción de marketing sobre un cliente, toma anotaciones sobre la conversación. El estudio de estas transcripciones mediante técnicas analíticas pueden dar información sobre la propensión de un cliente a la contratación de una cuenta nómina.

Los pasos realizados fueron:

- Eliminación de stopwords, es decir, palabras comunes como artículos, determinantes, verbos comunes que no dan información
- Palabras más frecuentes (unigramas) ordenadas según su frecuencia de aparición
- Agrupaciones de dos y tres palabras consecutivas (bigramas y trigramas)
- Análisis de los comentarios para detectar las connotaciones positivas y negativas
- Construcción de una nueva variable que indique según la información obtenida a través de text mining si el cliente es propenso o no



Caso 4: Propensión de contratación de cuenta nómina

Analítica

Text Mining

BIGRAMAS

no quiere inversión delegada

no interesado ffpp no

no necesita valores no

momento no cliente no

ahora no mando mail

cuenta nómina quedo disposición

traspaso valores agradece info

agradece llamada envio mail

traspaso ffpp ofrezco traspaso

no interesada no ffpp

necesita financiación no nomina

semana viene saldo cuenta

nomina no

TRIGRAMAS

no necesita financiación traspaso valores no momento no necesita cliente ahora no ahora no interesado traspaso ffpp no cliente no interesado momento no quiere cuenta nómina remunerada no interesado momento traspaso planes no posibilidad rentabilizar capital presento gestor distancia ofrezco traspaso ffpp préstamo preautorizado cliente disponible momento no traspaso planes cliente necesita financiación agradece preautorizado disponible momento de inversión delegada ofrezco posibilidad rentabilizar financiación agradece info no muestra interés

ahora no necesita ofrezco traspaso valores no quiere cambiar no ffpp fuera cuenta no nómina presento tutor distancia no interesado contratar traspaso pp no no quiere riesgo preautorizado no necesita momento no interesado cuenta nómina no ahora no quiere no valores fuera traspaso ffpp bonifi tutor distancia mando no quiere invertir informo traspaso pp distancia mando mail cuenta nómina cliente depositos cliente ahora no opera valores trasp ffpp no inversión delegada cliente ofrece ipf meses no opera acciones



Caso 4: Propensión de contratación de cuenta nómina

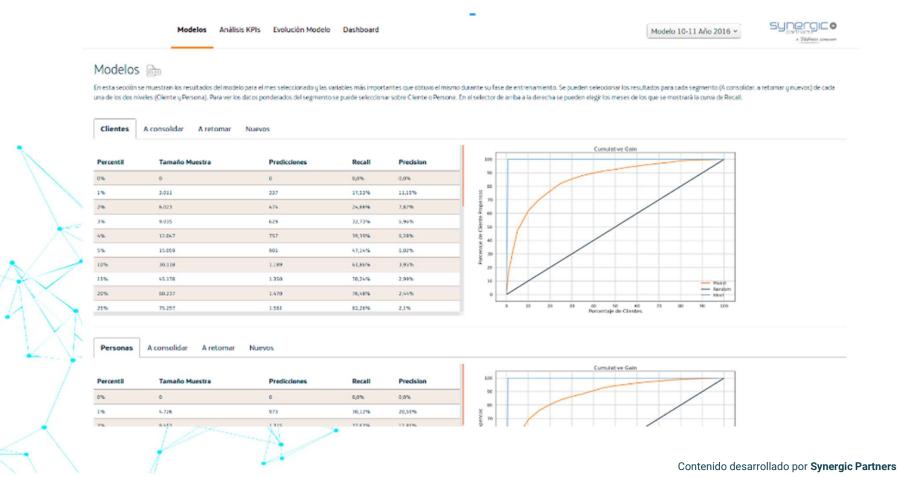
Resumen del proceso



Caso 4: Propensión de contratación de cuenta nómina

Output

Los resultados se mostraron mediante un dashboard interactivo que mostraba los resultados de los modelos, su evolución y el porcentaje de clientes clasificados como propensos frente al porcentaje de clientes que habían contratado la cuenta nómina..



Caso 4: Propensión de contratación de cuenta nómina

Output

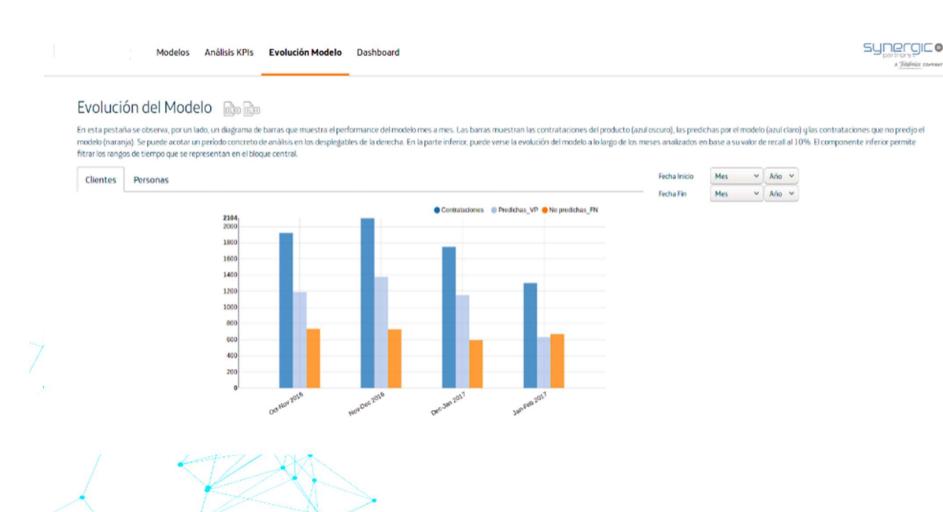






Caso 4: Propensión de contratación de cuenta nómina

Output





Upselling basado en DPI



El **objetivo** de este caso de uso es la **clasificación de los clientes** de la empresa partiendo del **consumo de datos de internet (DPI) realizado** por los mismos para, posteriormente, realizar campañas de marketing personalizadas.



Este caso de uso es solicitado por el departamento de **Business Intelligence** al detectar la **necesidad de conocer las** aplicaciones y páginas más visitadas por los clientes para así optimizar las futuras acciones comerciales.



El departamento de Bl detectó que el mercado actual se encuentra en un estado maduro y para poder penetrarlo hace falta un mayor conocimiento del cliente. Además se cuentan con 2 billones de registros web diarios por lo que se pueden obtener resultados significativos.



- → Valor empresa: Ingresos adicionales generados por las campañas personalizadas a nivel de cliente (*Upselling*).
- → Valor Cliente: Anticipación a futuras necesidades gracias al estudio del DPI, gustos y necesidades.





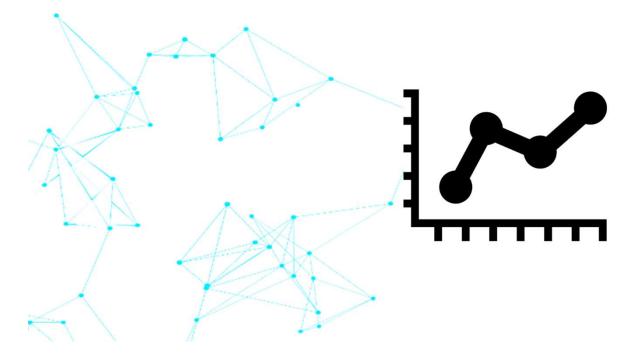
USE CASE: UPSELLING BASADO EN DPI

Upselling basado en DPI

Descripción

Hay que plantearse las siguientes cuestiones:

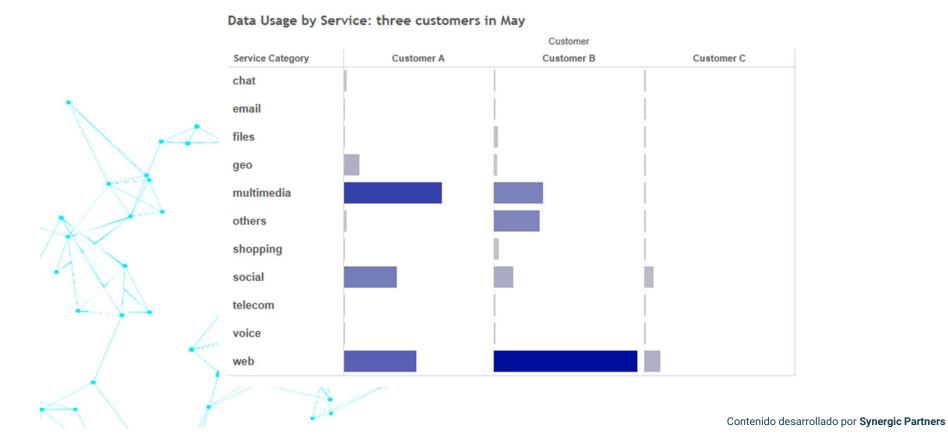
- ¿Cuál es el objetivo?
- ¿Qué tipos de datos tendríamos?
- ¿Qué tipo de problema analítico es?
- ¿Cómo lo solucionamos?



Upselling basado en DPI

Descripción

Imaginemos tres clientes, A,B y C. Los clientes A y B consumen ambos 2GB al mes, sin embargo podemos ver que el uso de estos datos es diferente. Por otro lado el cliente C no consume casi datos. Esto muestra que no podemos tratar a todos los clientes por igual



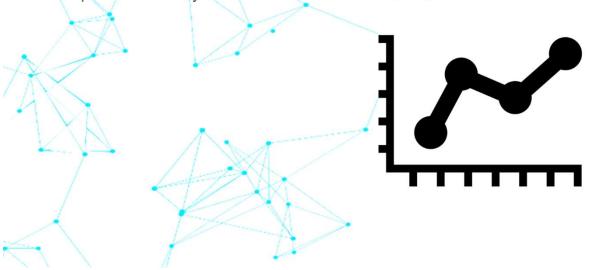
Upselling basado en DPI

Descripción

El principal objetivo del presente caso de uso es realizar campañas personalizadas.

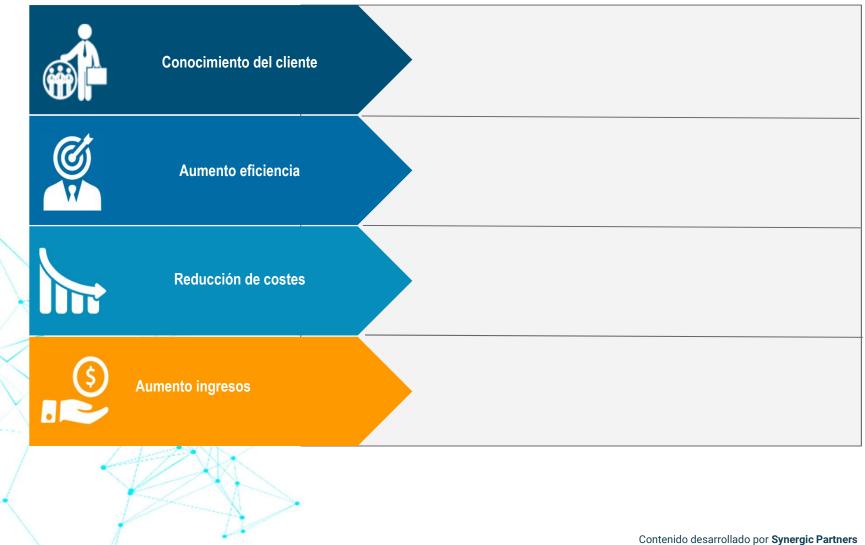
Para desarrollar el caso de uso se toman los datos de navegación de aproximadamente 6 millones de clientes de Telefónica. Una vez analizados se procede a realizar una microsegmentación (clasificación) de los mismos en base a unas categorías preestablecidas: aplicaciones que más usan, minutos de llamadas...

En una última fase y con la ayuda de un **algoritmo de agrupamiento** (clustering) se establecen **grupos de clientes:** los que llaman más y utilizan menos internet, los que utilizan más aplicaciones de mensajería, los que utilizan mayoritariamente redes sociales...



Upselling basado en DPI

Valor de Negocio



Upselling basado en DPI

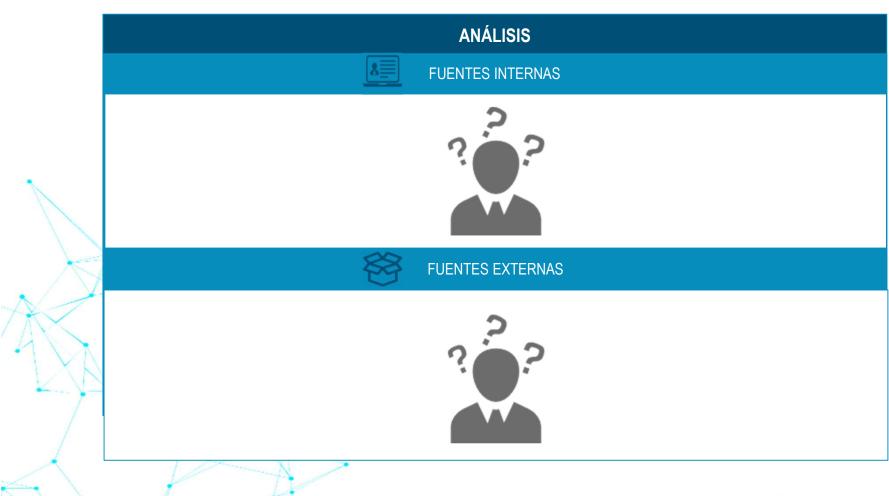
Valor de Negocio:



Upselling basado en DPI

Fuentes de datos

¿Qué fuentes de datos utilizarían para la realización del caso de uso?



Upselling basado en DPI

Fuentes de datos



FUENTES INTERNAS

- **Datos DPI:** uso de datos por usuario (agregados), día de uso, red (2G, 3G, 4G). Uso de datos en las diferentes Apps (Facebook, Youtube, Twitter,...), y tipo de terminal.
- Datos de consumo mensual clientes (agregados): minutos voz consumidos total, total de KB consumidos, número de SMS consumidos.
- Datos demográficos: edad, género, tiempo de permanencia en la compañía e índice de compra entre otros.



Upselling basado en DPI

Tratamiento de datos

Imaginemos todas las redes sociales, aplicaciones, datos de navegación... ¿cuántas aplicaciones tienes en tu teléfono móvil?

¿Puede dar lugar esto un problema en cuanto al volumen de datos y variables?

¿Cómo podría solucionarse?

Upselling basado en DPI

Tratamiento de datos

Se realizó un **proceso ETL** basado en la construcción de tablas de entrenamiento para los modelos, aplicando transformaciones a las tablas originales, uniendo tablas y derivando nuevas columnas de las ya existentes.

uniendo tablas y derivando nuevas columnas de las ya existentes.			
Los datos originales tenían mucha granularidad ya que venían dados por servicio. Estos se agruparon en 12 categorías como podemos ver en la tabla.			

Category	Serviceid		
	Apple iMessage	Yahoo Messenger	
ľ	Google Talk	IRC	
	Google Talk Data	Jabber	
Chat	Google Talk Push Notifications	Jabber SSL	
Chat	Google Talk Video	WeChat	
	ICQ	LINE	
	WhatsApp	Snapchat	
	Windows Live Messenger	Other_Realtime_Communication	
Email	Business_Email	Other_Email	
Files	Dropbox	Google_Drive	
	Kindle	Other_Network_Storage	
	iCloud		
C	EA Games Flat	Zynga	
Games	Candy Crush Saga	Other_Gaming	
Geo	Google Earth	Google Maps	
	Shazam	Deezer	
Multimedia	Spotify	Google_Music	
	Youtube	iTunes_Streaming	

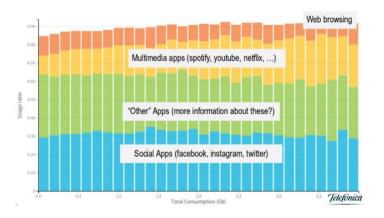
	Zattoo	last.fm	
	Napster (Rhapsody)	PANDORA_Radio	
	Amazon_Video	SoundCloud	
	Netflix	Tuneln_Radio	
	Sky_Go	Other_Realtime_Entertainment	
Others	Speedtest	Others	
Shopping	Amazon Mobile	Bitcoin	
Shopping	eBay	PayPal	
	Facebook	Pinterest	
Social	Foursquare	Twitter	
Social	Google+	Vine	
	Instagram	LinkedIn	
	Mein o2	Drillisch	
Talanam	Telekom	O2_Third_Brands	
Telecom	Vodafone	Other_Competitors	
	1und1		
	Skype	FaceTime	
	WeChat_Voice	SIP	
Voice	Viber	Tango_Voice	
voice	WhatsApp_Call	Blackberry_Voice	
	Viber_Call	Tango	
	Facebook_Voice	Yahoo_Messenger_Call	
Web	Google Search	Tor	
vved	Wikipedia	Other_Web_Browsing	

Upselling basado en DPI

Tratamiento de datos

- Como resultado de esta agregación se formaron
 48 variables nuevas, 4 para cada una de las categorías creadas:
- **category_total_bytes**: número total de bytes consumidos en esta categoría por un único cliente
- **category_row_count:** número de registros (filas) que son agregadas
- category_activity_time: uso total en segundos por cliente para dicha categoría
- category_service_count: número de aplicaciones diferentes usadas en esta categoría por cliente

Esta agregación por categoría permite analizar, por ejemplo, qué tipo de aplicaciones son responsables de un mayor consumo:



Upselling basado en DPI

Analítica

¿Cómo lo resolvemos?

Tenemos que plantearnos las siguientes cuestiones:

• ¿Qué tipo de problema analítico es?

• ¿Qué posibles soluciones hay?



Upselling basado en DPI

Analítica

¿Cómo lo resolvemos?

- ¿Qué tipo de problema analítico es? Clustering
- ¿Qué posibles soluciones hay? Algoritmo K-Means

La analítica principal consistió en dos fases:

- Segmentar los clientes mediante un algoritmo de clustering.
- Se aplicaron series temporales para ayudar a segmentar el comportamiento de los clientes respecto al uso de datos, estudiando la evolución temporal de este comportamiento.

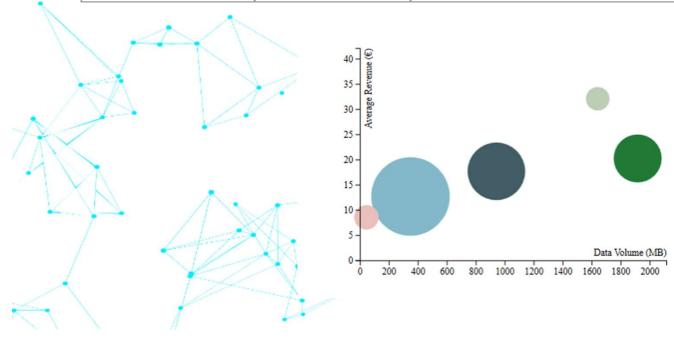


Upselling basado en DPI

Analítica

Clustering

Analítica	Algoritmo	Funcionamiento
Clustering	K-Means	El objetivo es encontrar patrones en el comportamiento de los clientes a nivel del uso de datos, del tiempo invertido en llamadas, SMS enviados, etc. El resultado fueron 5 clusters.

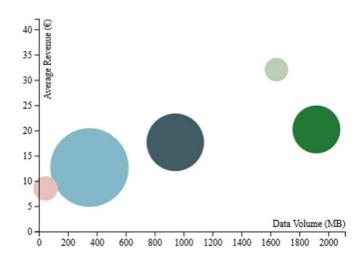


Upselling basado en DPI

Analítica

Clustering

¿Puedes identificar los clientes que forman los distintos clusters?

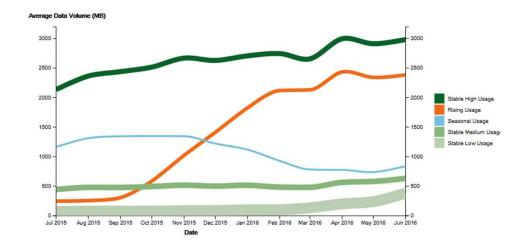


SOCIALMENTE ACTIVOS	COMPLETAMENTE ACTIVOS	MULTIMEDIA FANS	SÓLO ONLINE	PEQUEÑOS WHATSAPPERS
 Son los que pasan más tiempo online Tienen el mayor número de apps Los que más usan apps de chat Los que más usan apps sociales 	 Alto número de apps y tiempo online Participan mucho en apps de chat Usan mucho apps multimedia Poco uso de apps sociales 	 Su uso de internet se destina principalment e a películas, música, archivos Uso bajo de apps sociales 	 Pasan un tiempo muy limitado online Tienen las aplicaciones justas Utilizan mucho las apps sociales 	 No utilizan casi apps No pasan casi tiempo online No utilizan prácticamente apps sociales y multimedia Utilizan vagamente apps de chat

Upselling basado en DPI

Analítica

Series Temporales



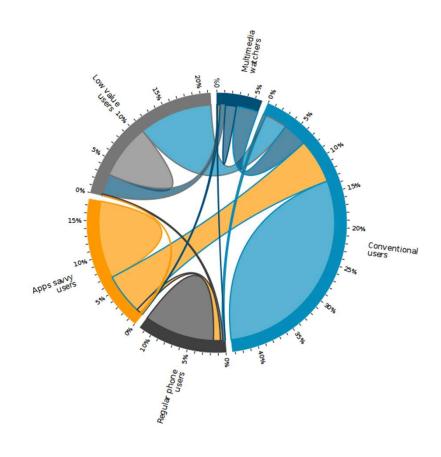
Analítica	Algoritmo	Funcionamiento
Series temporales	SAX	El modelo se basaba en el uso de datos mensuales de cada cliente durante un año. El objetivo es clusterizar los clientes en este caso según series temporales sobre el uso de datos a lo largo del año. El resultado fueron también 5 clusters.

Stable High Usage	Stable Medium Usage	Stable Low Usage	Rising Usage	Seasonal Usage
Description: 11% of customers Highest ARPU: 29 € Youngest average age Disproportionately male (59%) Highest postpaid share (93%)	Description: • 13% of customers • ARPU: 21 €	Description: • 66% of customers • Lowest ARPU: 12 € • Oldest average age • Highest prepaid & value brand share	Description: • 7% of customers • High ARPU: 28 € • High postpaid share (90%)	Description: • 4% of customers • High ARPU: 24 €

Upselling basado en DPI

Analítica

Segmentación Dinámica



_	Analítica	Algoritmo	Funcionamiento
	Segmentación dinámica	K-Means	Se ejecutó el algoritmo de clustering, una vez conocido el número de clusters, en diferentes meses para ver el movimiento del número de clientes de un cluster a otro en diferentes momentos de tiempo.

Upselling basado en DPI

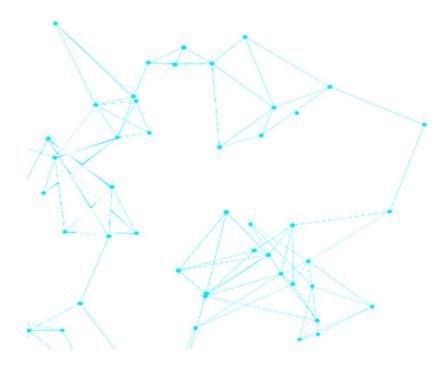
Resumen del proceso



Upselling basado en DPI

Output

El conjunto de resultados obtenidos se mostraron mediante un formato web en un workshop en Alemania. Este workshop fue impartido por el departamento técnico a los departamentos de Marketing y Ventas.



Upselling basado en DPI

Resultados

Inicialmente se conocía que 10 Millones de clientes no usaban datos de internet, lo cual suponía un *ARPU* de unos 3€.

Al segmentar a los clientes se obtienen las siguientes conclusiones de la segmentación:

- Active Allroudners: El mayor ARPU (29€) corresponde a los clientes que concentran la mayor cantidad de aplicaciones, tanto de chat como multimedia. No hacen gran uso de apps sociales.
 También son aquellos que hacen más uso de llamadas por voz. La media de edad es joven.
- Reluctant WhatsAppers: Aquel segmento con menor ARPU (9€) corresponde a los clientes de mayor edad donde los contratos predominantes son prepago.
- El 7% de los clientes pasa de no usar datos de internet a usar 2KB. El 90% son de contrato postpago.
- ─ El 66% de los clientes corresponde a aquellos de mayor edad y poco uso de internet.
- Se encontraron 3 grupos: los clientes que no usan datos, los que usan por primera vez y los que tienen un uso de datos estable.
- Los clientes que no usan datos y los que los usan por primera vez tienen un ARPU bajo, excepto el grupo Active Allroudners.

Este conocimiento tan granular de los clientes permite enfocar las campañas de marketing a aquello que el cliente realmente quiere o necesita.

