



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**OPTIMIZACIÓN EN LA NEGOCIACIÓN DE CAMPAÑAS PUBLICITARIAS EN
INTERNET**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

CRISTÓBAL ELÍAS HAYE LEÓN

PROFESOR GUÍA:

RODOLFO URRUTIA URIBE

MIEMBROS DE LA COMISIÓN:

PATRICIO CONCA KEHL

NICOLÁS FRITIS COFRÈ

SANTIAGO DE CHILE

2016

**RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE:** Ingeniero Civil Industrial

POR: Cristóbal Elías Haye León

AÑO: 2016

PROF. GUÍA: Rodolfo Urrutia Uribe

**OPTIMIZACIÓN EN LA NEGOCIACIÓN DE CAMPAÑAS PUBLICITARIAS EN
INTERNET**

La publicidad toma una innumerable cantidad de formas y tamaños, pasando por medios convencionales como la calle y la televisión. Hoy en día existe una fuerte tendencia en el crecimiento de publicidad en la internet por diversas razones, en particular porque es un medio que llega a una gran masa de personas con un costo relativo por persona mucho menor que la televisión o prensa escrita.

La relevancia del impacto del internet en la publicidad se ve fuertemente reflejada en un aumento desde 0% en el 2002 hasta un 10% en el 2014 en el gasto total en publicidad en Chile, llegando a órdenes de magnitud de 2.000 MM de pesos chilenos. Es por esta razón que nace la necesidad de optimizar la manera de invertir el dinero con el fin de tener mejores resultados en las campañas publicitarias.

El cliente del trabajo de título es una empresa que provee el servicio de compra programática, la cual consiste en la compra de espacios publicitarios en la web, a los clientes de la empresa de publicidad Grupo Havas, que estén interesados en hacer avisaje a través de internet. La empresa *Affiperf* implementa, para estos clientes, campañas publicitarias en internet a través de plataformas diseñadas para participar en la compra en tiempo real de espacios publicitarios obteniendo un margen en la compra.

Las campañas publicitarias en la web tienen la particularidad de que es posible medir el rendimiento de las campañas en base a distintas métricas, entre las cuales están el costo por mil impresiones (CPM), costo por realizar un clic (CPC) y el ratio clics e impresiones (CTR). Por esto, cada avisaje tiene un objetivo concreto basado en estos indicadores los cuales pretenden distribuir de la mejor manera posible la compra de espacios publicitarios en internet para tener un mejor retorno sobre la inversión en base a estas métricas.

Los costos incurridos en la elaboración de sus servicios es significativo y no siempre se logran los objetivos de las campañas debido a que asume la variabilidad del comportamiento del usuario de internet. Esto da pie al caso de estudio, el cual consiste en proponer un modelo de regresión basado en los datos históricos de la empresa proveedora del servicio de compra programática, para optimizar negociaciones sobre las métricas de rendimiento en las campañas publicitarias con los avisadores.

Los resultados del modelo propuesto presentan una mejor predicción del CTR de un 3,8% por sobre lo realizado por *Affiperf* lo que tiene un valor estimado de 124,3 [UM] en la negociación para la realización de campañas publicitarias. Por otro lado, la predicción del CPC no obtiene resultados debido al alto nivel de variabilidad del indicador.

Tabla de Contenido

Capítulo 1. Introducción	1
1.1. Antecedentes	1
1.2. Descripción del Proyecto	2
1.2.1. Los Agentes	2
1.2.2. El Cliente	3
1.2.3. El Proceso	4
1.2.4 Configuración de Campañas Publicitarias.....	7
1.2.5 El Modelo de Negocio	10
Capitulo 2. Justificación y Metodologías	14
2.1. Justificación	14
2.1.1 Variabilidad del Retorno	14
2.1.2 Negociación de <i>Affiperf</i> con sus Clientes	16
2.1.3 Los Riesgos del Negocio de Compra Programática.....	16
2.2. Objetivos	18
2.2.1. Objetivo General	18
2.2.2. Objetivos Específicos	18
2.3. Marco Conceptual.....	19
2.3.1 Transformación y Validación de Atributos	19
2.3.2 Métodos de Segmentación.....	21
2.3.3 Evaluación de medidas de desempeño de modelos.	21
2.3.4 Descripción Creatividades.....	24
2.3.5 Herramientas Computacionales	27
2.4 Metodología	28
2.4.1 Análisis, Transformación y Selección de Atributos.....	28
2.4.2 Segmentación de Campañas	29
2.4.3 Análisis de Elasticidades.....	29
2.4.4 Modelo Propuesto	29
Capitulo 3. Análisis, Transformación y Selección de Atributos.....	31
3.1 Base de Datos	31
3.1.1. Descripción de los Datos.....	31
3.2 Generación de Atributos	33
3.3 Validación de Atributos	35

3.3 Segmentación de Estrategias Publicitarias	37
3.4 Supuesto Cualitativo	40
Capítulo 4. Valoración Económica de la Predicción	42
4.1 Valorización de los Clientes	43
4.2 Probabilidad de Deserción de un Cliente	46
4.2 Costos de la Predicción	47
Capítulo 5. Evaluación de Modelos Predictivos	50
5.1 Predicción de CTR.....	50
5.1.1 Resultados sin Segmentación.....	50
5.1.2 Resultados con Segmentación.....	52
5.2 Predicción de Costo por Clic (CPC).....	55
5.2.1 Resultados sin Segmentación.....	55
5.2.2 Resultados con Segmentación.....	57
Capítulo 6. Conclusiones.....	59
6.1 Conclusiones de la Metodología	59
6.2 Recomendaciones a la Empresa	62
6.3 Trabajo Futuro	62
Glosario.....	63
Bibliografía.....	64
Anexo y Apéndices	66
Anexo 1: Clusters Creatividades.....	66
Anexo 2: Resultados Regresión sin Segmentación CTR.....	67
Anexo 3: Resultados Regresión sin Segmentación CPC.....	67

Índice de Tablas

Tabla 1, Ejemplos Clientes <i>Affiperf</i>	11
Tabla 2, Ejemplo Reporte para el Cliente de <i>Affiperf</i> , Resultados de la Campaña Publicitaria.....	16
Tabla 3, Tipos de Errores de <i>Affiperf</i> , Costo Relativo y Probabilidad.....	17
Tabla 4, Transformaciones a Distintos Tipos de Atributos	19
Tabla 5, Resumen de Análisis de Componentes Principales	20
Tabla 6, Resumen K-Means.....	21
Tabla 7, Distintos Rendimiento de Creatividades.....	26
Tabla 8, Agrupación de Creatividades Según Criterio Davies Bouldin.....	34
Tabla 9, Representación Visual de Segmentos de Creatividades.....	34
Tabla 10, Cantidad de Elementos en los Segmentos de Creatividades.....	35
Tabla 11, Resultado Análisis de Componentes Principales	37
Tabla 12, Agrupación de Creatividades Según Criterio Davies Bouldin.....	38
Tabla 13, Resultado Segmentación Estrategias Publicitarias	39
Tabla 14, Esquema de Proyección de la Negociación de las Campañas Publicitarias ..	42
Tabla 15, Sobreestimación de las Métricas de Rendimiento de las Campañas Publicitarias Reales vs Modelo.....	51
Tabla 16 Sobre Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo.....	51
Tabla 17, Sub Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo.....	52
Tabla 18, Sobre Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo.....	53
Tabla 19, Sub Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo.....	55
Tabla 201, Sobre Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo.....	56
Tabla 2221, Sub Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo.....	57
Tabla 2322, Sobre Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo.....	57
Tabla 2523, Resumen de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo.....	61

Índice de Ilustraciones

Ilustración 1, Elaboración de plan de medios. Elaboración propia	2
Ilustración 2, División de venta publicitaria en Internet. Elaboración propia.....	2
Ilustración 3, Compra programática. Fuente: Affiperf 2015, Conferencia de RTB, Chile..	4
Ilustración 4, Protocolo RTB. Fuente: Affiperf 2015, Conferencia de RTB, Chile.	5
Ilustración 5, Protocolo RTB. Fuente: Affiperf 2015, Conferencia de RTB, Chile.	6
Ilustración 6, Desglose del margen de la compra en RTB, Elaboración Propia	11
Ilustración 7, Variabilidad del Margen Diario	14
Ilustración 8, Tasa de Efectividad.....	15
Ilustración 9, Matriz de confusión. Fawcett, Tom (2006). "An Introduction to ROC Analysis"	22
Ilustración 10, Tamaños Creatividades	24
Ilustración 11, Creatividades campaña Dettol	25
Ilustración 12, Esquema de la Metodología a Implementar.....	28
Ilustración 13, Análisis de componentes principales	36
Ilustración 14, Distinta Utilidades Según Clientes	43
Ilustración 15, Valor Presenta a Diferentes Tasas Mensual.....	45
Ilustración 16, Valor Presente, Distintos Periodos Mensuales	46
Ilustración 17, Número de Campañas	47
Ilustración 18, Separación datos Entrenamiento y Prueba.....	50

Capítulo 1. Introducción

1.1. Antecedentes

Hoy en día con la globalización y el crecimiento poblacional, es necesario utilizar medios de comunicación para transmitir a los segmentos de interés la idea o producto que se está ofreciendo, con el fin de ser conocido y competitivo en mercados cada día más grandes y complejos. La publicidad en internet es una tendencia hoy en día llegando a ser un 10%^[1] del total de los gastos en publicidad en Chile.

La penetración del internet en Chile alcanza el 66,4%^[2] de la población actual, llegando al rededor de 11 millones de habitantes, por lo que una de las mayores cualidades de internet es que llega a una fuerte y creciente masa de personas en un corto periodo de tiempo, por lo que es un excelente medio para hacer conocida la campaña a promocionar. Además de que los usuarios son aquellos potenciales clientes que tienen una relativa capacidad de adquisición e interés en función de los sitios en los que navega, por ejemplo los usuarios cotizando automóviles en sitios de distribuidoras.

Dentro de la logística en la cual se expone la publicidad, existe un número no menor de variables de decisión que el publicista debe tomar para elaborar las campañas web. Hay firmas que se dedican a optimizar las campañas en base a información capturada en internet a través de *cookies*¹ y datos propios del individuo proporcionados por el mismo. Cuando no se utiliza este tipo de programas debido a su elevado costo o por desconocimiento, la labor del publicista se complejiza y debe recurrir a su experiencia para optimizar las campañas.

El objetivo es generar mejores retornos sobre la inversión de publicidades utilizando evidencia estadística en base a los resultados históricos, con el fin de argumentar mejoras en el proceso de implementación de campañas en base al protocolo RTB, el cual consiste en el protocolo de venta del espacio publicitario en base a una subasta en tiempo real para los interesados en la compra del mismo. El resultado es valorizar las diferencias entre el estado actual y el estado propuesto. La empresa permitirá poner en práctica las metodologías bajo la supervisión de un tutor y los datos con un pacto de confidencialidad.

“Programmatic trading is like teenage sex – everyone is talking about it, everyone thinks everybody else is doing it when they’re not, and those who are doing it probably aren’t doing it very well,” Google’s VP Matt Britti, 11 July, 2014.

¹ Información adherida al navegador del usuario, la cual permite entre otras cosas, monitorear la navegación de la persona.

1.2. Descripción del Proyecto

1.2.1. Los Agentes

El proceso de avisaje en internet en base al protocolo de subasta en tiempo real se desglosa en la interacción de cuatro entidades separadas. El proceso comienza con la intención de un avisador en realizar una campaña publicitaria. El primer agente provee el mensaje en forma de *creatividades*, junto con un presupuesto y un público objetivo al cual desea impactar con la campaña. La segunda entidad es quien elabora un *plan de medios*, la cual puede dividirse en tres grandes medios: prensa escrita y televisión, la calle e internet como se ilustra en la ilustración 1.

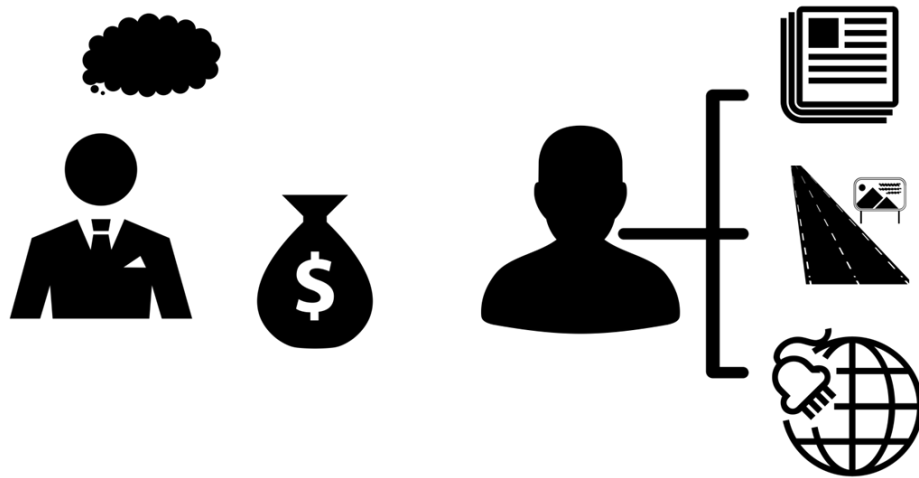


Ilustración 1, Elaboración de plan de medios. Elaboración propia

La compra de espacios publicitarios en internet, se divide en tres grandes aristas como lo muestra la ilustración 2.

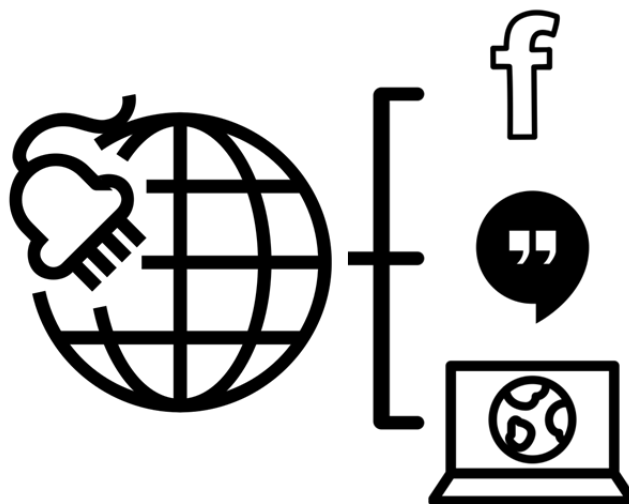


Ilustración 2, División de venta publicitaria en Internet. Elaboración propia.

Esta división ocurre debido a la gran participación de mercado de los dos primeros del esquema Facebook y Google, venden sus espacios publicitarios por un medio diferente y no entra en el caso de estudio. El enfoque del análisis está en la tercera entidad la cual es, el conjunto de todos los sitios que proveen un espacio publicitario para la venta según el protocolo de *RTB* o subasta en tiempo real.

El tercer agente es la empresa proveedora del servicio de *compra programática*. Esta empresa compra espacios publicitarios en internet, basado en el protocolo de compra en tiempo real (*RTB*) a través de plataformas comerciales (*Mediamath, Appnexus, StrikeAd* entre otras) que tienen algoritmos integrados para subastar por espacios publicitarios disponibles. Este proceso será detallado más adelante.

El cuarto agente es el usuario final, el cual navega por la web y es impactado por la publicidad. Este agente tiene un comportamiento aleatorio desde la perspectiva de *Affiperf* y los avisadores, debido a que actúan según los intereses personales y son parte del caso de estudio.

1.2.2. El Cliente

El cliente de la memoria es la empresa que provee el servicio de compra programática. *Affiperf* es una empresa creada en Francia en el momento en que Google incorpora un nuevo método para hacer publicidad el cual es la subasta en tiempo real (*RTB*). *Affiperf* funciona como el punto de conexión entre los avisadores y las plataformas que demandan espacios publicitarios (*DSP*), puesto a que implementa las campañas en las mismas plataformas.

En Chile la incorporación se realizó el año 2012 y ha tenido una creciente participación en los planes de medios. En otras palabras muchos planes de publicidad que anteriormente se reducían en la prensa escrita, televisión y publicidad en la calle comenzaron a asignar presupuesto a la compra programática.

Previo a la realización de las campañas publicitarias, se realiza un proceso de negociación sobre las métricas de rendimiento de las mismas la que actualmente se realiza en base a la experiencia del encargado en relaciones con los avisadores en la empresa *Affiperf*. El proceso presenta dos diferentes tipos de problemas, el primero es subestimar las métricas de rendimiento de las campañas previo a la realización, con lo que no se logra el acuerdo establecido con el cliente. El segundo problema es sobre estimar los resultados, lo que genera un mejor retorno sobre la inversión para el cliente pero condiciona la imagen de *Affiperf* para futuras campañas.

El presente estudio se enfoca en establecer modelos con fundamento teórico para reducir estos dos tipos de problemas, minimizando los costos en los que se incurre y optimizando la utilidad para *Affiperf*.

1.2.3. El Proceso

El proceso de compra programática es la manera de participar en el protocolo de subasta en tiempo real, utilizando para esto información histórica, análisis del comportamiento en internet y experiencia entre muchos otros factores como se esquema en la ilustración 3, la cual combina el protocolo RTB junto con distintas fuentes de información y análisis extraídos de la web.



Ilustración 3, Compra programática. Fuente: Affiperf 2015, Conferencia de RTB, Chile.

Es natural pensar que actuar en el protocolo no puede ser un proceso manual debido a la rapidez que requiere y a la inmensa cantidad de veces que se realiza por segundo. Es por esto que se elaboran plataformas automáticas que participan en el protocolo. Estas plataformas demandan espacios publicitarios, por lo que se les llama "plataformas del lado de la demanda", las cuales interactúan con las plataformas en el lado de la oferta. Las plataformas en el lado de la oferta agrupan múltiples espacios publicitarios de diferentes sitios para venderlos.

El protocolo de RTB es cuando la oferta se une con la demanda, como lo ilustra la figura 4.

Lado de la Oferta

Lado de la Demanda

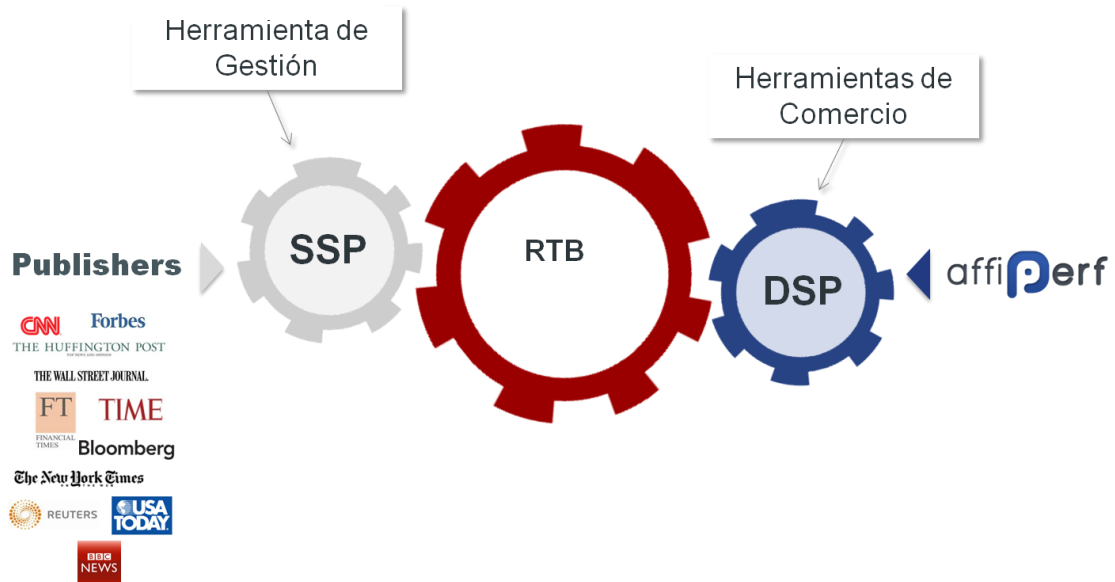


Ilustración 4, Protocolo RTB. Fuente: Affiperf 2015, Conferencia de RTB, Chile.

Existen una gran variedad de plataformas del lado de la demanda, los cuales tienen diferentes atributos y características especiales. Dentro de ellos, la empresa Affiperf utiliza los siguientes *DSP* para realización de campañas publicitarias:

1. StrikeAd
2. Appnexus
3. Mediamath
4. Doubleclick Bid Manager

Los *DSP* tienen las siguientes características comunes entre ellas:

- Acceso rápido a los datos
- Datos en tiempo real
- Categorías estandarizadas

El estudio se basará únicamente en el *DSP* Mediamath por las siguientes razones:

- Utilizado en 80% de campañas publicitarias para display.
- Campañas duraderas en el tiempo
- Alto nivel de inversión
- Focalizado en computador personal.

1.2.3.1 Protocolo de subasta en tiempo real RTB

El protocolo de subasta en tiempo real puede ser descrito en 5 pasos como lo muestra la ilustración 5:



Ilustración 5, Protocolo RTB. Fuente: Affiperf 2015, Conferencia de RTB, Chile.

1. Un usuario ingresa a un sitio en el cual hay un espacio publicitario disponible para el uso del protocolo, este se inicia.
2. El espacio disponible se anuncia por parte de las plataformas del lado de la oferta SSP a todas las plataformas del lado de la demanda a la que estén asociados.
3. Cada plataforma del lado de la demanda evalúa según la configuración que se le integra según las campañas publicitarias si decide subastar por el espacio y cuánto dinero está dispuesto a invertir.
4. Cada DSP envía una oferta por el espacio (la oferta es cero si no hay interés por la compra).
5. El mayor oferente gana la impresión y paga la subasta a segundo precio más un centavo, es decir paga la oferta del segundo mayor postor.

El proyecto estará enfocado en el tercer paso del protocolo, el cual depende de la configuración de la campaña publicitaria por parte de la empresa proveedora del servicio de compra programática. En los DSP se configuran las campañas publicitarias en términos de varias variables de decisión para lograr los objetivos que se plantean en

cada una de ellas. Las variables de decisión varían entre los diferentes *DSP*, pero en *Mediamath* se pueden desglosar como se expone en el siguiente punto.

1.2.4 Configuración de Campañas Publicitarias

Para entender cómo se configuran las campañas publicitarias, es necesario entender los diferentes objetivos que se pueden proponer en cada una y las condiciones que se establecen en ellas. Junto con esto, se desglosarán las diferentes decisiones que debe tomar la empresa de compra programática para poder llevar a cabo la campaña y cumplir con los objetivos.

1.2.4.1 Objetivos de las Campañas Publicitarias

Los objetivos de las campañas publicitarias se pueden dividir en tres para el caso de estudio:

1. Tráfico en el sitio: se desea aumentar la cantidad de visitas al sitio publicitado en el cual se encuentra el producto o servicio que se ofrece. También se utiliza para promocionar una idea. Éste objetivo se traduce en el número de clics sobre la creatividad que se realizan durante la campaña publicitaria, puesto que ellos refieren al sitio en promoción.
2. Reconocimiento de Marca: se busca dar a conocer el producto a algún segmento de la población pero no necesariamente tráfico en el sitio. Se pretende incrementar el conocimiento de la campaña. Este tipo de objetivo se traduce en un número de impresiones dentro del segmento objetivo
3. Búsqueda de Acciones: la intención es que un número determinado de usuarios realice una acción determinada dentro del sitio avisado en la campaña publicitaria. Esta acción puede ser llenar un formulario, realizar una compra o alguna otra opción determinada por el avisador. Estas campañas se miden por el número de las acciones y se miden con un contador en el sitio que se despliega luego de haberla realizado. En general el contador está en el despliegue de una pantalla que dice gracias por realizar tal acción.

1.2.4.2 Variables de Decisión en la Programación de Campañas Publicitarias

Cada campaña se configura con la intención de lograr de mejor manera el objetivo de la misma. Las variables se desglosan de la siguiente manera:

1. Presupuesto: cantidad de dinero especificado para la realización de la campaña publicitaria.
2. Máximo pago por CPM: Máxima disposición a pagar por mil impresiones.
3. Frecuencia de impresiones por usuario: número máximo de impresiones que puede recibir un usuario por unidad de tiempo. En general se manejan las impresiones diarias y las impresiones semanales. Esta variable se determina a través de cookies en el navegador del usuario y solo se puede realizar en aquellos clientes que permiten su uso en el navegador.
4. Gasto por unidad de tiempo: la cantidad de dinero máximo a utilizar dentro de un intervalo de tiempo, en general se manejan a nivel diario y a nivel mensual.
5. Distribución del presupuesto: se puede distribuir el presupuesto de dos maneras según el gasto por unidad de tiempo, los cuales son gasto parejo durante el periodo o lo antes posible hasta lograr el gasto esperado.
6. Segmentación: cada *DSP* provee segmentación de los sitios según la similitud entre ellos, Mediamath provee segmentos de sitios los cuales son determinados en base algoritmos que recorren los sitios buscando palabras clave. La segmentación aumenta el costo de la compra debido a que agrega el valor de tener una referencia en los sitios en los que se pretende aparecer para realizar la campaña.
7. Capacidad de entrega: se puede establecer un máximo de impresiones en alguna unidad de tiempo.
8. Seguridad de la marca: es un bloqueo para evitar aparecer en sitios de mala reputación y mala visibilidad de los espacios publicitarios.

Las variables de decisión descritas son el objeto de estudio, puesto que se desea conocer cuánto afecta cada una de ellas en los objetivos de las campañas.

1.2.4.3 Tipos de Campañas y Estrategias Publicitarias

Cada campaña publicitaria cuenta con diferentes objetivos según las métricas de rendimiento que se proyectan para ella. A grandes rasgos se pueden separar las campañas en 3 tipos:

- Campañas Promocionales: Son aquellas que promocionan un producto o idea con corta duración en el mercado. Por ejemplo campañas publicitarias para la realización de maratones o venta de productos en el Ciber Monday².
- Campañas de Reconocimiento de Marca: Son campañas que son persistentes en el tiempo y buscan recordar a los clientes la existencia de la marca. Se realiza en productos que se venden por largos periodos de tiempo, como por ejemplo CocaCola.
- Campañas de Re marketing: Son campañas que buscan impactar a usuarios que tuvieron alguna actividad con el producto a vender. Un ejemplo es buscar a los clientes que realizaron una cotización de autos en los últimos 3 meses para impactarlos, con el fin de aumentar la probabilidad de compra.

Cada campaña publicitaria puede contener una o más estrategias por parte de *affiperf* las cuales son las que se aplican al momento de participar en el protocolo en tiempo real. Estas estrategias también se pueden separar en 3 grandes grupos los cuales son los siguientes:

- Estrategias de Alcance: Son estrategias las cuales buscan impactar a la mayor cantidad de usuarios sin incurrir en costos de segmentación aplicados por las plataformas (DSP). El objetivo es llegar a nuevos usuarios que navegan en sitios aleatorios de internet.
- Estrategias Contextuales: Son estrategias que buscan impactar a usuarios en sitios que la plataforma segmenta según su contenido. Un ejemplo es intentar comprar espacios publicitarios en sitios catalogados como sitios de automóviles, para publicitar un producto asociado a los autos.
- Estrategias de Re marketing: Se impactan a usuarios que han interactuado con sitios específicos a través de *cookies*. Un ejemplo es guardar todos los usuarios que han entrado al sitio de cotizaciones de automóviles de Citroen, para luego impactarlos en los otros sitios en los cuales navega y tienen a la venta espacios publicitarios para RTB.

² Día dedicado a la venta de productos mayoritariamente electrónicos con grandes descuentos.

1.2.5 El Modelo de Negocio

La empresa *Affiperf* provee el servicio de compra programática utilizando diferentes DSP para esto. El modelo de negocio consiste en proponer un plan publicitario al cliente, incluyendo un presupuesto, un costo y una meta y el retorno se obtiene al realizar el mismo objetivo a un menor costo. A continuación se desglosan los componentes principales del negocio de la compra programática.

La principal ventaja competitiva de la empresa es que debido a su actividad a nivel global y su asociación con *Havas Media*, que es una de las empresas de publicidad más grande del mundo [10]. Es por esto que tiene una gran cartera de clientes alrededor del mundo, luego el nivel de inversión que se maneja permite lograr economías de escala y por lo tanto realizar el negocio a un costo competitivo frente a sus competidores. Por lo tanto el acceso a la cartera de clientes de *Havas Media* es su mayor ventaja competitiva, puesto que con ello logra obtener mejores tratos con las plataformas del lado de la demanda, reduciendo el margen que obtienen ellas (las plataformas de lado de la demanda) al prestar sus servicios para realizar la compra programática, con el fin de obtener los altos montos de inversión que *Affiperf* obtiene de los clientes de *Havas Media*. Entendiendo esta ventaja competitiva, *Affiperf* puede proveer el servicio a un precio menor que el de la competencia sin perjudicar el margen al realizar el servicio.

Por otra parte, *Affiperf* ha sido pionero en el ingreso a nuevos países como Chile, al cual llegó el 2012, innovando en nuevos países convirtiéndose en la empresa símbolo en el ejercicio de compra programática. Este valor agregado se impone en el valor de la marca y la confianza que entrega a sus clientes alrededor de 220 países con más de 2 trillones de impresiones disponibles mensuales [11].

1.2.5.1 Segmento de Mercado

El segmento de mercado manejado por la empresa son todas aquellas firmas asociadas al grupo *Havas Media*, el cual consiste en más de 1.000 marcas a nivel mundial. En Chile actualmente hay grandes firmas de la empresa *Havas Media* que activamente compran servicios de *Affiperf*, en la siguiente tabla se muestran algunos de los clientes con los que se han realizado inversiones en el último año:

Peugeot	RedBull	Gasco
DIRECTV	Watts	LG
Philips	Cruz Verde	Ideal
Natura	LAN	Citroen

Tabla 1, Ejemplos Clientes *Affiperf*

En el presente año 2015 *Affiperf* ha trabajado con más de de 44 diferentes marcas en Chile y más de 1.000 campañas publicitarias de los mismos en el rubro de la compra programática.

1.2.5.2 Alianzas Estratégicas

La mayor alianza estratégica de *Affiperf* es la unión como único proveedor de compra programática para la empresa de publicidad *Havas Media*. Esta alianza le provee los clientes, la inversión y acceso a una posición favorable para establecer contratos con sus proveedores.

La empresa tiene alianzas vitales para el negocio, las cuales son con las plataformas del lado de la demanda (DSP), debido a que en estas plataformas se implementan las campañas publicitarias y se tiene acceso al protocolo de venta de espacios publicitarios en internet (o RTB, real time bidding sus siglas en inglés). Las plataformas permiten participar en la compra de espacios web para el avisaje y cobran un porcentaje de la inversión por el servicio prestado. Como ejemplo en la ilustración 6 se muestra un desglose del costo de comprar un espacio publicitario (comprar una impresión), separado por los diferente agentes involucrados:

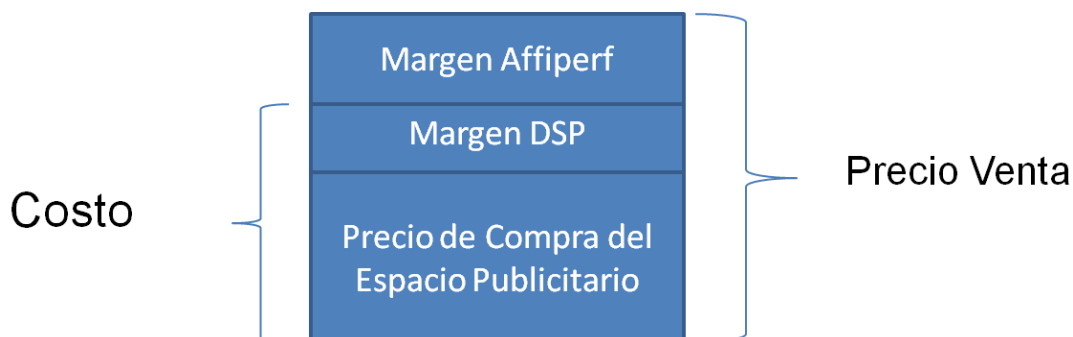


Ilustración 6, Desglose del margen de la compra en RTB, Elaboración Propia

El primer elemento del precio de compra en la figura, representa el precio de venta del espacio publicitario en la subasta del sitio en el cual se compró el inventario (espacio publicitario a la venta para el protocolo RTB), el segundo es el margen que entrega la plataforma del lado de la demanda por permitir implementar campañas publicitarias en ella y acceder al protocolo RTB y finalmente el margen de *Affiperf*. La suma de estos tres costos es lo que se le cobra a la empresa que está realizando el avisaje, y el costo es el valor de la compra agregado al margen que obtiene el DSP.

La alianza con el DSP permite reducir el margen que obtiene, con el fin de mantener los niveles de inversión que pasan por *Affiperf*, luego puede vender el servicio a un menor costo y/o obtener un mayor margen. Esta ventaja competitiva es vital en el negocio y lo mantiene como el mayor proveedor de servicio de compra programática en Chile.

1.2.5.3 Propuesta de Valor

El servicio provee la posibilidad realizar publicidad en internet para los usuarios del mismo. Internet tiene un nivel de penetración del 66,4% en Chile actualmente, lo cual lo convierte en un mercado atractivo para quienes desean impactar a estos potenciales clientes. El método para impactar usuarios permite integrar diferentes tipos de información con 3 diferentes objetivos.

El primero es identificar en qué tipos de sitios los usuarios serán impactados con el fin de hacer publicidad coherente con el usuario. Como ejemplo se pueden publicitar productos electrónicos en sitios que están categorizados como noticias de información tecnológica (esta categorización es realizada por proveedores de los DSP).

El segundo objetivo es poder medir en tiempo real a cuántos usuarios ha impactado una campaña publicitaria, cómo han reaccionado los usuarios ante la publicidad y calcular diferentes métricas de retorno sobre la inversión. Se puede ilustrar calculando cuánto cuesta que un usuario realice un clic en la publicidad, considerando el clic como un ingreso al sitio que se desea publicitar como por ejemplo en campañas de la Teletón se evalúa cuál es el tráfico adquirido por la campaña publicitaria.

El tercer objetivo es llegar a una masa importante de usuarios a un costo relativamente bajo, en una tecnología que está en constante crecimiento y desarrollo el cual es el internet. Cada vez hay más usuarios de internet y menos usuarios de otros medios como la prensa escrita y la televisión. Esto se explica por una tendencia al interés por la novedad tecnológica y la globalización.

Luego el valor agregado, además del servicio de compra programática, por parte de *Affiperf* se desglosa de la siguiente manera [11]:

- Cobertura Global: Tiene experiencia en mercados internacionales y acceso a compra de espacios publicitarios en todo el mundo.
- Acceso a Inventario Exclusivo: Por su alto nivel de inversión, tiene acceso a sitios de internet con espacios premium de publicidad.

- Verificación de Publicidad y Seguridad de Marca: Provee alianzas estratégicas con empresas que verifican la publicidad con el fin de que sea apta para RTB, además de agregar filtros de sitios considerados como ofensivos o que podrían afectar negativamente la marca en promoción.
- Confidencialidad de Datos: Garantiza confidencialidad de datos frente a sus competidores.
- Segmentación Avanzada: Alianzas con proveedores de información que categorizan los sitios, utilizando palabras claves para ellos.
- Análisis e Integración de Datos: Activamente utilizan la data obtenida en tiempo real para optimizar las campañas publicitarias para mejorar métricas de retorno.
- Accesos a Datos Históricos de Rendimientos: Permite utilizar datos proveídos por terceras empresas para garantizar mejores resultados de las campañas publicitarias.

Capítulo 2. Justificación y Metodologías

2.1. Justificación

La justificación del proyecto se puede dividir en dos áreas que son complementarias entre sí. La primera área es la justificación económica de las campañas publicitarias realizadas por la empresa. Este punto consiste en la validar si la campaña propuesta generará utilidades a la empresa.

La segunda justificación es la propia factibilidad de lograr las metas de las campañas publicitarias en función de las métricas de rendimiento negociadas con el avisador.

2.1.1 Variabilidad del Retorno

Affiperf se encuentra al final de la cadena de elaboración de campañas publicitarias en internet. La empresa enfrenta el riesgo de lograr o no las campañas con el presupuesto asignado y por lo tanto los márgenes obtenidos son erráticos, cómo se muestra en la ilustración 6, en donde el eje de las abscisas corresponde a la fecha en la que comienzan las campañas publicitarias y en el eje de las ordenadas se ilustra el margen obtenido en las campañas que comienzan en esa fecha.

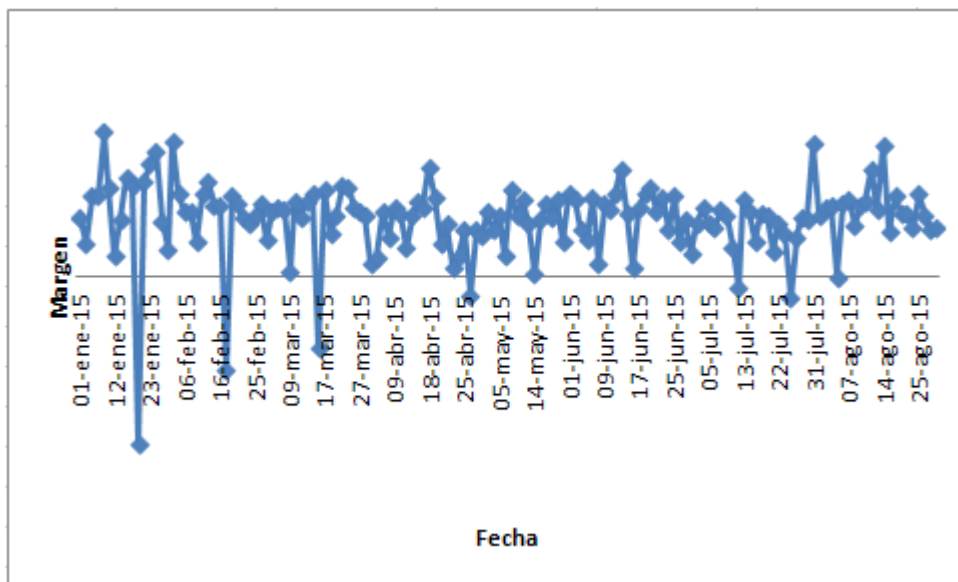


Ilustración 7, Variabilidad del Margen Diario

Debido al contrato de confidencialidad, los montos no están explícitos. La variabilidad se puede explicar principalmente por dos fenómenos.

1. Errores en la implementación de las campañas, lo que genera costos más elevados de lo negociado incurriendo en pérdidas para lograr la meta de los clientes con el fin de retenerlos. Un error común (0,8% de los casos) es la programación de una campaña publicitaria con un presupuesto mayor al que el cliente establece, con lo cual se gasta más dinero del que se obtiene por la venta, generando pérdidas.
2. Comportamiento de los usuarios de internet en los que campañas con objetivos específicos no se logran. Un ejemplo es realizar la venta de **N** clics a un precio **X** y obtener en la práctica de la realización de la campaña un costo **X+ε** por cada clic, generando pérdidas de **N*ε**.

El enfoque del proyecto es disminuir el segundo problema con evidencia estadística y que la empresa logre un mayor y más sustentable margen. La ilustración 8 muestra la tasa de efectividad de las campañas publicitarias.

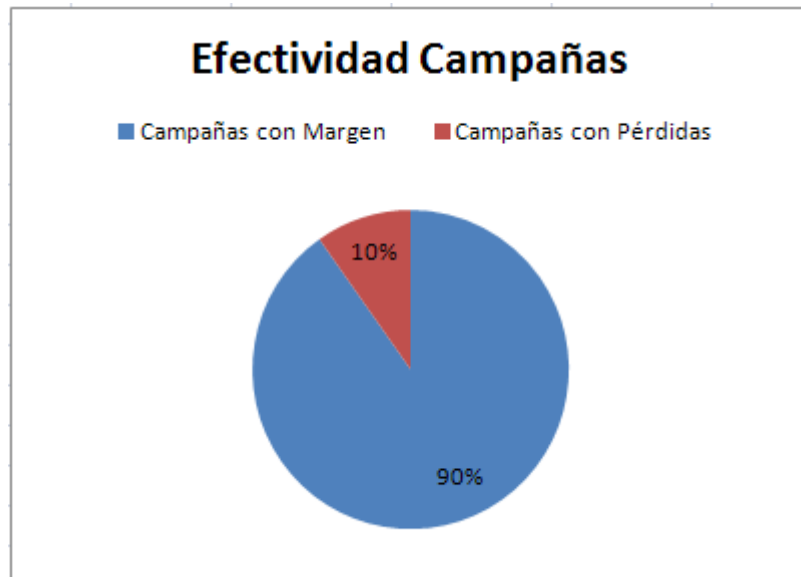


Ilustración 8, Tasa de Efectividad

Basado en los últimos 8 meses, condiciones son favorables para *Affiperf* pero existe holgura para mejorar los rendimientos y por lo tanto los retornos de la empresa.

2.1.2 Negociación de *Affiperf* con sus Clientes

Affiperf acuerda objetivos explícitos para cada campaña publicitaria con el avisador. Se establecen un objetivo primario y otros objetivos secundarios, en las campañas de publicidad en internet. Una vez negociado y aceptado el plan, comienza la implementación de las campañas publicitarias y finalmente se comparan los resultados con las proyecciones. En la siguiente tabla se muestra un ejemplo de una campaña publicitaria que no logró el acuerdo establecido:

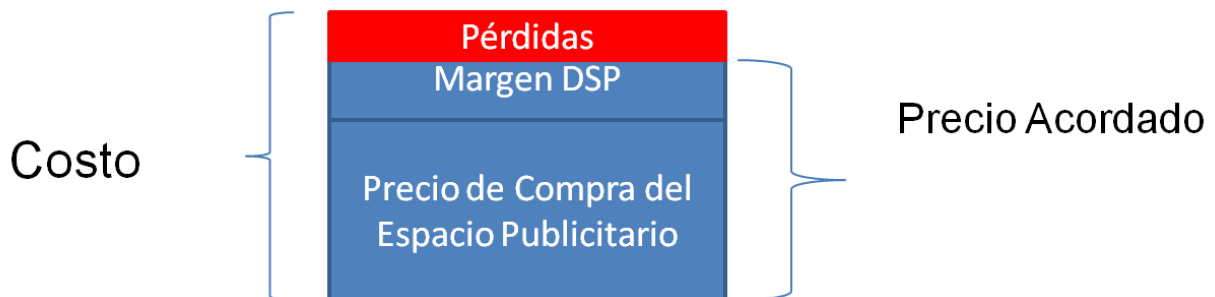
	Impresiones	Clics	CTR	CPC [UM]
Proyectado	570.484	570	0,11	459
Obtenido	613.073	516	0,08	659
Diferencia Porcentual	+7,4%	-10,4%	-28,2%	+21,3%

Tabla 2, Ejemplo Reporte para el Cliente de *Affiperf*, Resultados de la Campaña Publicitaria

La imagen representa las campañas publicitarias en compra programática en la primera fila. Se puede observar que se cumplieron las proyecciones de impresiones (+7,4%), pero no se logró el nivel de clics (-10,4%) y tampoco se lograron las predicciones de CTR (-28,2%) ni las de CPC (costo superior del 21,3%) . Este resultado no implica que *Affiperf* no genere margen, sino que no cumplió las proyecciones que se acordaron en un principio debido a que la negociación en base a la experiencia estuvo lejos de la realidad. Es el perfecto ejemplo sobre el caso de estudio actual.

2.1.3 Los Riesgos del Negocio de Compra Programática

Affiperf tiene dos fuentes de incertidumbre sobre las utilidades que obtiene con las campañas publicitarias. El primer riesgo es el que incurre es acordar un precio por unidad (ya sea CPC o CPM) y obtener un precio mayor a la hora de comprar (error de tipo 1). La siguiente ilustración muestra el caso:



En donde el costo de comprar el espacio publicitario es mayor que a lo que se acordó vender. Esto puede ocurrir con campañas que dependan del comportamiento del

usuario final (como por ejemplo clics), en donde los usuarios impactados no reaccionan al nivel que se especula. El resultado de esta situación lleva a un no cumplimiento de la campaña por que se agota el presupuesto con anterioridad o incurrir en pérdidas para lograr la campaña.

La segunda fuente de incertidumbre (error de tipo 2) es el no lograr la campaña publicitaria debido a falta de inventario, baja reactividad de los usuarios o errores humanos. Aunque se logre una unidad de compra a un menor precio de la venta, es posible que las proyecciones sean más altas de lo que se logre. En este escenario aunque *Affiperf* pueda obtener un margen incurre en el riesgo de perder al cliente para futuras campañas debido a no cumplir sus proyecciones. Este costo se avalúa como el valor presente del cliente a lo largo del ciclo de vida del mismo.

Finalmente el error de tipo 2 es mucho más costoso que el error de tipo 1 pero compensa la probabilidad de ocurrencia. La siguiente tabla caracteriza los errores y los costos relativos:

	Costo	Probabilidad
Error de tipo 1	Mediano o Bajo	Media
Error de tipo 2	Muy Elevado	Baja o Media

Tabla 3, Tipos de Errores de *Affiperf*, Costo Relativo y Probabilidad

A lo largo del presente informe se intentarán caracterizar las probabilidades y costos de ocurrencias del error de tipo 2.

2.2. Objetivos

2.2.1. Objetivo General

Establecer una metodología con fundamento teórico para optimizar la negociación de las campañas publicitarias basadas en el protocolo RTB entre *Affiperf* y sus clientes, basado en información histórica

2.2.2. Objetivos Específicos

1. Identificar variables de decisión al momento de establecer una campaña.
2. Asociar campañas que tienen comportamientos comunes (clusters).
3. Utilizar análisis estadísticos para establecer elasticidades de los rendimientos de las campañas publicitarias en función de sus variables de decisión, compararlos y evaluarlos
4. Finalizar estableciendo una metodología genérica para los distintos tipos de campañas.

2.3. Marco Conceptual

Para la aplicación de la metodología descrita, es necesario explicar modelos y algoritmos que se utilizarán en cada paso.

2.3.1 Transformación y Validación de Atributos

Las transformaciones de atributos dependen del tipo del mismo, a continuación se presenta una tabla con las respectivas transformaciones válidas para los diferentes tipos de atributos [5].

Escala	Significado	Operación permitida
Nominal	Nombre de objetos	Comparación
Ordinal	Orden total entre atributos	Transformación monótona
Intervalo	El cero y la unidad arbitrario	$F(x)=a+bx, a>0$
Proporcional	Dado el cero, unidad arbitraria	$F(x)=bx$
Absoluta	Dado el cero y la unidad	$F(x)=x$

Tabla 4, Transformaciones a Distintos Tipos de Atributos

Estas transformaciones se realizaron en base a la experiencia y al conocimiento del negocio, además de agregar algunas otras arbitrarias que podrían explicar los datos.

Para la validación de atributos se utilizó el análisis de componentes principales. Este método consiste en describir la varianza de la nube de datos en vectores ortogonales entre sí, ordenándolos por nivel de descripción de la varianza de los datos. El método consta de obtener la matriz de varianzas y covarianzas de los atributos. Esta matriz $X \in R^{M \times M}$ en donde M es la dimensión de los atributos a considerar, en la diagonal X_{ii} tiene las varianzas de cada atributo y en los elementos restantes $X_{ij}, i \neq j$ tiene las covarianzas de los atributos i, j . Esta matriz representa la variabilidad de cada atributo con respecto al otro, por lo que la idea detrás del análisis de componentes principales es caracterizar la varianza de los datos con la menor cantidad de vectores posibles.

El método calcula los valores propios de la matriz, los cuales son los vectores que cumplen con la siguiente propiedad. Sea v un vector de dimensión M , v es un vector propio de la matriz X si cumple con la siguiente propiedad: $Xv = \lambda v$ con λ un

escalar entero. Se puede demostrar que todos los vectores propios de una matriz son ortogonales y que la magnitud del valor propio describe cuanta dispersión hay en la matriz X es la dirección del valor propio. Luego cada vector propio describe una dimensión de la matriz de varianzas y covarianzas, ordenadas según el valor propio asociado a cada vector propio. El método ordena los vectores según la magnitud del vector propio y selecciona aquellos vectores que describen la mayor parte de la matriz de varianzas y covarianzas deteniéndose bajo algún umbral aceptable de varianza descrita por cada vector propio. Entonces el método queda descrito de la siguiente manera:

PCA (Análisis de componentes principales)

1	Obtener matriz de varianzas y covarianzas
2	Obtener valores y vectores propios de la matriz
3	Ordenar los vectores según la magnitud del valor propio asociado
4	Seleccionar atributos en el orden obtenido en 3, hasta cumplir algún criterio para detenerse.
5	Aplicar al modelo los atributos seleccionados

Tabla 5, Resumen de Análisis de Componentes Principales

El principal objetivo es reducir la dimensionalidad de los atributos, justificando que vectores explican de mejor manera los datos.

2.3.2 Métodos de Segmentación

En el proyecto se elaboran dos diferentes tipos de segmentación los cuales se planean comparar en base a los resultados finales. El primer tipo de segmentación es en base al algoritmo K-Means, el cual es un modelo no supervisado e independiente de la intuición del realizador de la segmentación. A continuación se muestra el pseudocódigo del algoritmo K-Means [6]:

K-Means

1	Establecer K puntos dentro de los datos y fijarlos como centroides
2	Asignar cada dato a su centroide más cercano en base a la distancia euclidiana, asignando a un segmento
3	Recalcular los centroides como el promedio de todos los puntos asignados a cada segmento
4	Realizar paso 2 y 3 hasta que no cambien los centroides entre iteraciones

Tabla 6, Resumen K-Means

Por otro lado el criterio Davies Bouldin establece una proporción entre que tan cerrados o compactos son los clusters con la distancia entre los centroides. El indicador se obtiene de la siguiente manera:

$$DB_{nc} = \frac{1}{nc} \sum R_i$$

$$R_i = \text{Max}_{i \neq j} R_{ij}$$

$$R_{ij} = \frac{S_i + S_j}{d_{ij}}$$

En donde S_i es la distancia promedio de los elementos de un cluster a su centro, d_{ij} es la distancia euclidiana entre los centros de los clusters i y j y nc es el número de clusters. Mientras menos sea este indicador, los clusters son más compactos y están a una mayor distancia que bajo algunos criterios es una forma de comparar distintos tipos de segmentaciones [7].

2.3.3 Evaluación de medidas de desempeño de modelos.

Para evaluar un modelo predictivo es necesario establecer los costos en los que se incurre con cada tipo de error y el beneficio obtenido por aciertos en el modelo. Los modelos de clasificación separan dos tipos de observaciones, las cuales para diferenciarlas se considerarán como positivos y negativos y se les asignará un valor +1 y -1 respectivamente. Para clarificar la situación se muestra en la ilustración 1 la matriz de confusión [1].

	p' (Predicted)	n' (Predicted)
P (Actual)	True Positive	False Negative
n (Actual)	False Positive	True Negative

Ilustración 9, Matriz de confusión. Fawcett, Tom (2006). "An Introduction to ROC Analysis"

La matriz de confusión compara el resultado real de los datos en el set de entrenamiento con respecto a lo predicho por el modelo de clasificación. Los elementos de la matriz se enuncian a continuación:

- True Positive: El modelo clasifica como positivo a un elemento del set de training que se sabe es positivo.
- False Positive: El modelo predice un valor positivo para el elemento cuando en realidad es negativo.
- False Negative: El modelo predice un valor negativo cuando en realidad el elemento es positivo.
- True Negative: El modelo predice un valor negativo a un elemento negativo.

En base a esta matriz se elaboran diferentes métricas de rendimiento que serán explicadas a continuación:

2.3.3.1 Accuracy

Accuracy se define como la proporción de predicciones que fueron acertadas (True Positive y True Negative) sobre la cantidad de observaciones [2], [3], [4]. la siguiente fórmula ilustra el cociente:

$$\frac{(TP + TN)}{(TP + FP + FN + TN)}$$

En donde TP se refiere a True Positive, TN es True Negative, FN es False Negative y FP es false positive. La interpretación de esta métrica es la proporción de veces que el modelo clasifica correctamente los valores dentro del set de validación (también se puede realizar el conjunto de datos de entrenamiento).

La métrica es válida de utilizar para tener una intuición de que tan calibrado es el modelo y si está *sobre ajustado* en el set de entrenamiento, comparando el accuracy de los datos de entrenamiento y los de validación, junto con evaluar la calidad del modelo únicamente según los aciertos. La métrica es poco representativa cuando el costo/beneficio de cada elemento de la matriz difiere, en casos donde cometer un error o acierto genera más o menos utilidad. En este caso el accuracy no se considera una métrica representativa del valor que entrega el modelo. Por otro lado esta medida es poco significativa cuando la proporción de valores positivos y negativos está muy desequilibrada. Consideremos un caso donde son 1.000 observaciones de las cuales 995 son negativas, luego un clasificador que seleccione todas las predicciones como negativas obtendrá un accuracy del 99,5%, lo que se contradice con la validez del modelo.

2.3.3.1 Precision

Precisión se define como la proporción de veces que se acertó a una predicción positiva. Es el cociente entre el número de aciertos positivos (TP) por sobre la suma de veces que se clasificó como positivo [2], [3], [4]. La siguiente fórmula ilustra el indicador:

$$\frac{TP}{TP + FP}$$

La interpretación intuitiva de la fórmula es el porcentaje de cuantas veces el modelo dice que es positivo y en la realidad lo es. Esta métrica es solo un indicador del porcentaje de aciertos del modelo al valor positivo, luego no considera costos/beneficios diferentes en la matriz de confusión. Además es invariante al número de valores negativos, por lo que no da una visión general del modelo sino que solo a los valores categorizados como positivos. Junto con ello, la métrica indica un mejor modelo mientras mayor sea este, pero no considera un sobre ajuste (si se habla de el indicador en los datos de entrenamiento). Es por esto que es necesario compararlo con otras métricas como Recall.

2.3.3.1 Recall

Recall se define como la proporción de casos positivos que fueron correctamente clasificados. En otras palabras es el cociente entre el número de aciertos positivos por sobre la suma de valores positivos en la base de datos [2], [3], [4]. La siguiente fórmula ilustra el indicador:

$$\frac{TP}{TP + FN}$$

La intuición detrás de esta métrica es calcular el número de valores positivos asignados sobre el total de valores positivos en la base de datos. Esta métrica es invariante ante los valores negativos de la base de datos y nuevamente no considera el costo/beneficio de cada error o acierto. Su función es generar una idea de cuál es el porcentaje de acierto en los casos positivos de la base de datos.

2.3.4 Descripción Creatividades

Cada campañas publicitaria tiene un conjunto de creatividades con las cuales se elabora la misma. En la siguiente ilustración se muestran algunos de los diferentes tamaños de creatividades posibles:

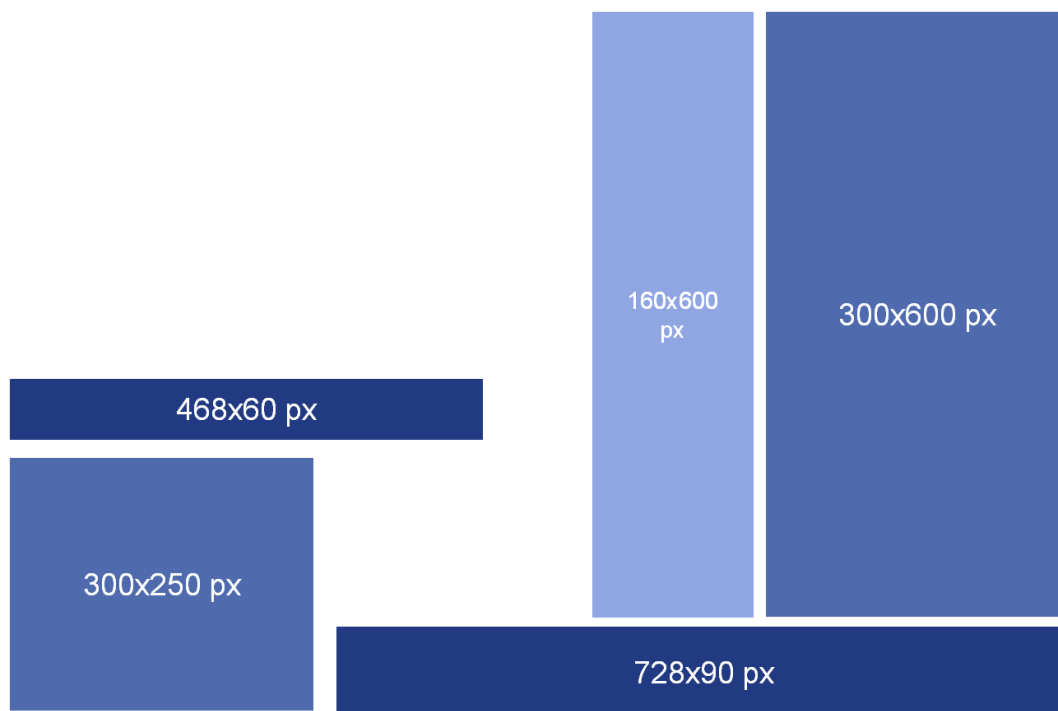


Ilustración 10, Tamaños Creatividades

en donde "px" se refiere a las dimensiones en píxeles³. Los tamaños de creatividades pueden alterar el resultado de las campañas debido a las posiciones que ocupan en los sitios los cuales pueden tener mejor visibilidad que otros. Para ilustrar esta situación se presenta la comparación de las creatividades de una campaña publicitaria de la marca Dettol, en la cual todas las creatividades tienen la misma secuencia de movimientos pero se muestran en diferentes etapas para ilustrar la situación.

³ Píxeles son las unidades básicas de tamaños de pantallas para los computadores tradicionales.

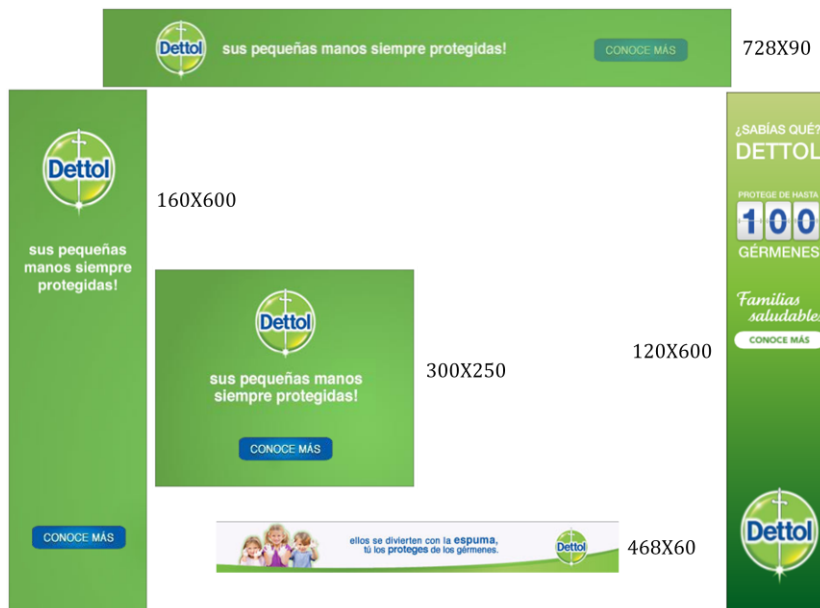
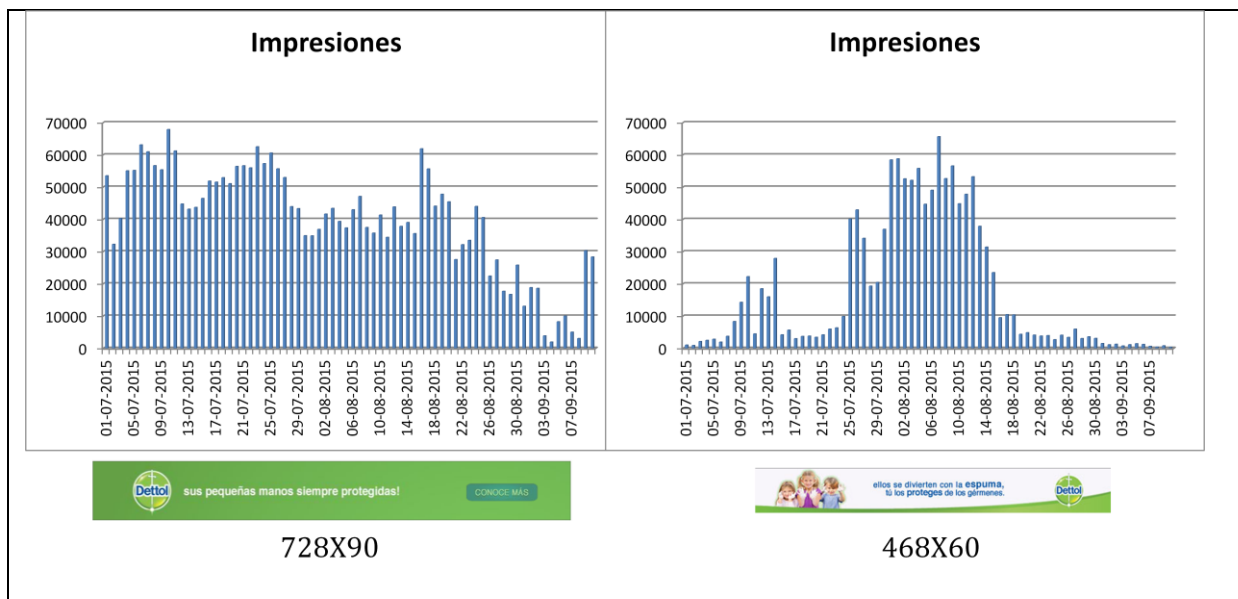


Ilustración 11, Creatividades campaña Dettol

Se puede observar que tiene seis diferentes creatividades de diferentes tamaños, por lo que pueden aparecer en diferentes lugares dentro de un sitios. Cada sitio provee inventario según los tamaños. En la siguiente ilustración se comparan las creatividades de 728x90 pixeles y la de 468x60 en términos de resultados de las campañas:



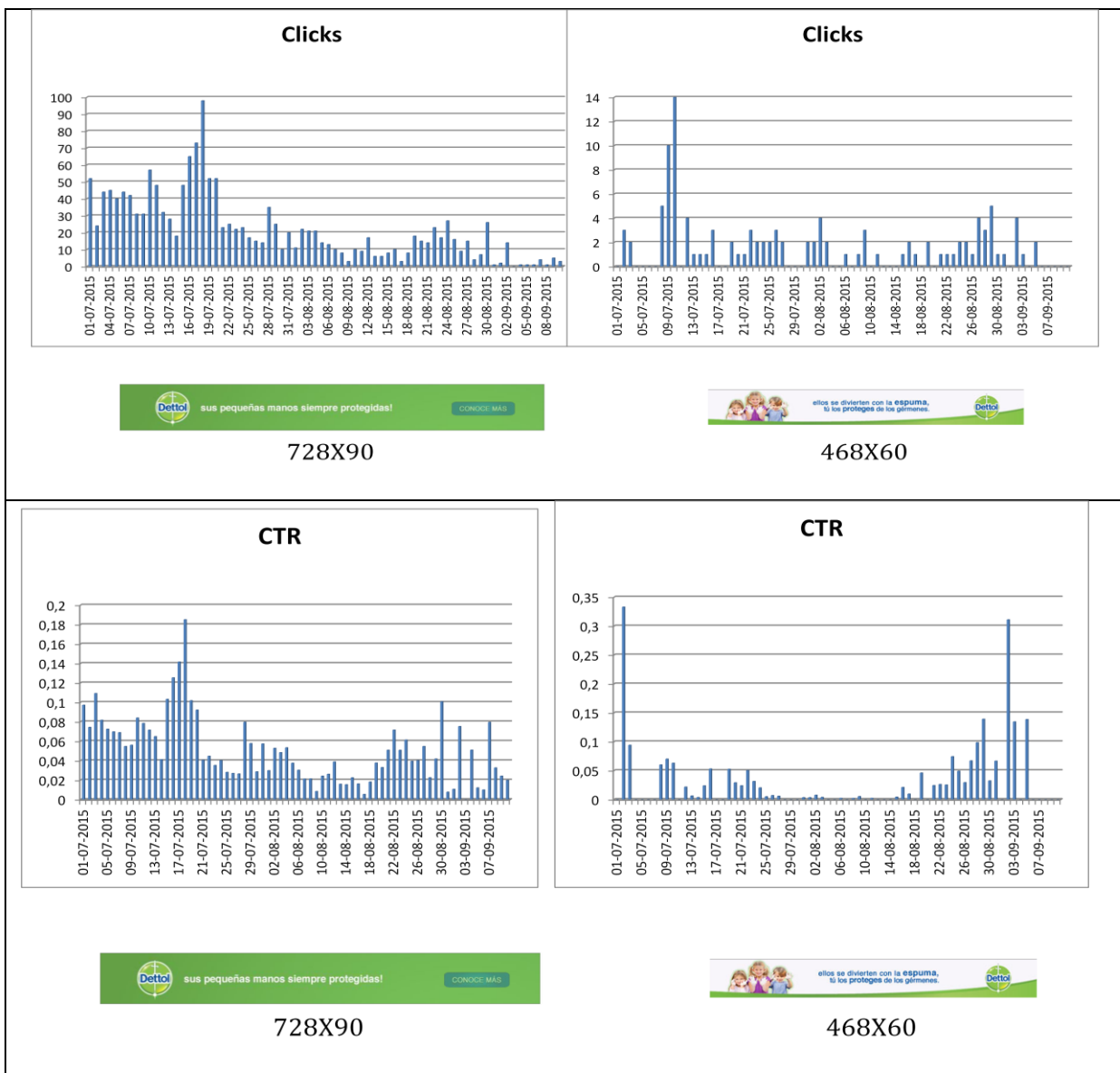


Tabla 7, Distintos Rendimiento de Creatividades

Consideremos el primer elemento de la tabla el cual ilustra la cantidad de impresiones diarias durante el primero de julio hasta el siete de septiembre. La creatividad de 728x90 tiene un mayor nivel de impresiones pero esto se puede deber a diferentes causas. La primera posible explicación es que hay más espacios para la venta de creatividades de esta dimensión por lo que se compra más de ella. La segunda explicación puede ser que las subastas para ese tamaño de creatividad tienen un menor precio por lo que la apuesta asociada a esa creatividad gana más subastas que la creatividad de 468x60.

El segundo elemento de la tabla muestra el número de clics en cada creatividad. Es natural pensar que la primera creatividad es más atractiva para los usuarios pero también se debe considerar que esta creatividad tubo más oportunidades de ser expuesta al usuario final, por lo que no es evidente cual de ambas creatividades tubo un mejor rendimiento.

Finalmente en la última fila de la tabla se comparan los rendimientos en la razón clics y creatividades. El valor es cuantos clics se obtienen por cada impresión. Luego aquí tampoco es evidente cual creatividad tiene mejor rendimiento absoluto por sobre la otra creatividad, por lo que se decide incluir los tamaños de creatividad como un atributo del modelo.

2.3.5 Herramientas Computacionales

Para la elaboración de los análisis estadísticos se utilizaron diferentes herramientas y softwares estadísticos para distintas etapas del proyecto. Para la realización de segmentaciones tanto de estrategias publicitarias como de creatividades se utilizó RapidMiner Studio 5.Ink. Para la aplicación de modelos estadísticos predictivos se utilizó R. Por otra parte para la modificación de atributos, generación de tablas y gráficos se usó Excel 2007.

2.4 Metodología

La metodología a implementar consiste en realizar un análisis exploratorio de las variables de decisión que se pueden modificar para establecer una estrategia de publicidad. Junto con esto analizar los resultados que se obtuvieron en base a las variables de decisión, para de esta manera tener una visión general del problema mismo.

Se consideran las campañas a nivel mensual, en las cuales se pretende realizar una segmentación en base al comportamiento agregado de las mismas, para obtener mejores predicciones sobre el rendimiento de las campañas de las que se realizan actualmente. La metodología propuesta se separa en cuatro componentes principales para elaborar un análisis profundo con respecto a cuanto afectan las decisiones tomadas a los diferentes tipos de campañas publicitarias, la ilustración 9 muestra el esquema a proceder de la metodología propuesta.

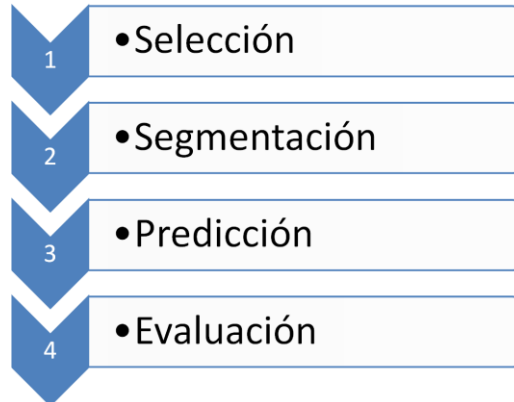


Ilustración 12, Esquema de la Metodología a Implementar

2.4.1 Análisis, Transformación y Selección de Atributos

Para comenzar se realizará un análisis exploratorio de los atributos, buscando cuales son lo más relevantes para explicar el rendimiento de las campañas publicitarias, basados en 2 objetivos de ella. El primer objetivo de la campaña a analizar es el costo por clic, lo que determina cuánto dinero se tuvo que invertir para que un usuario en internet tuviera una reacción positiva y decidiera hacer un clic en el aviso, la segunda métrica de rendimiento es la tasa de clics sobre las impresiones entregadas, es decir cuántas impresiones se necesitaron para que un usuario realice un clic.

Para la selección de atributos se elaboraron distintas transformaciones y combinaciones de ellos, buscando cuáles de ellos proveen un mejor nivel de explicación de los datos basado en un análisis de componentes principales. El análisis se utilizó debido a que reduce la dimensionalidad del problema, mejorando la capacidad de interpretación de los atributos que se mantienen.

Una vez realizado lo anterior, se agruparán los tipos de campañas publicitarias por segmentos, usando como criterio de distinción el comportamiento de la efectividad de las campañas en función de las variables de decisión (o atributos). En otras palabras se agruparan campañas publicitarias que se vean afectadas de maneras similares ante la misma decisión. El objetivo de hacer esta segmentación es poder dividir el problema en partes comunes, para luego abordar cada sub problema por separado usando la misma metodología a proponer como término de la memoria.

2.4.2 Segmentación de Campañas

Encontrados los atributos (o transformaciones de ellos) que expliquen la mayor variabilidad de los datos, se efectuará una segmentación de las campañas publicitarias basadas en estos mismos atributos. Para realizar esta segmentación se aplicarán dos diferentes métodos de clusterización siendo el primero el algoritmo K-Means debido a la facilidad de la implementación y a la necesidad de utilizar un método sin supervisión para compararlo con uno supervisado. E segundo un modelo aglomerativo jerárquico, el cual es supervisado y depende de la experiencia y conocimiento del negocio, el objetivo detrás es realizar una comparación entre las dos diferentes maneras de segmentar y comparar resultados finales para cada uno de los casos. Para evaluar el método no supervisado de segmentación se utilizará el indicador Davies Bouldin, el cual establece una métrica que relaciona que tan compactos son los segmentos con la distancia entre los centros de los mismos, en los datos.

2.4.3 Análisis de Elasticidades

Finalmente se propone un análisis de elasticidades de las variables de decisión dentro de cada uno de los segmentos propuestos para un nivel desagregado de las campañas publicitarias a nivel diario. Para la realización de este análisis se propone realizar modelos lineales con la variable dependiente siendo las tres diferentes métricas de rendimiento establecidos (CPC, CTR) y las variables independientes a los atributos obtenidos en la primera etapa. El objetivo cuantificar el nivel en el que afecta cada atributo (o transformación de él) en el resultado de los diferentes segmentos de las campañas publicitarias.

2.4.4 Modelo Propuesto

Se propone utilizar un modelo de regresión lineal dentro de los clusters que se generaron de las estrategias publicitarias, para obtener cuanto afecta cada variable de decisión a los diferentes indicadores de rendimiento de las campañas publicitarias. Este modelo se propone debido a que interesa saber cómo afectar de mejor manera el

rendimiento de las estrategias y cuantificar el efecto marginal de ellos en las campañas globales. La facilidad de implementación, explicación e interpretación son idóneos para la elaboración de reglas para diferentes conjuntos de estrategias y la facilidad de aplicación de la misma.

El objetivo es predecir el comportamiento de las campañas publicitarias para realizar una proyección más adecuada al momento de la negociación previa a la realización del avisaje en internet, para reducir los costos que estos producen.

Capítulo 3. Análisis, Transformación y Selección de Atributos

La elaboración del proyecto se divide en tres grandes áreas. La primera consiste en el levantamiento de datos, en el que se describe la obtención de datos, los diferentes niveles de agregación y el pre procesamiento de ellos con tal de obtener la estructura necesaria para realizar la metodología propuesta. La segunda etapa es el manejo, transformación y evaluación de los atributos con el fin de obtener las decisiones que explican los resultados en base a diferentes criterios de evaluación. La tercera es la aplicación de los modelos detallados en la metodología para obtener los resultados en base a los datos pre procesados y transformados. Finalmente en la última etapa se realizan interpretaciones de los resultados obtenidos.

3.1 Base de Datos

Los datos provienen de la misma plataforma de estudio la cual es *Mediamath*. Debido al contrato de confidencialidad y para mantener la misma escala de magnitud, todos los datos fueron normalizados en el intervalo 0-1 utilizando la siguiente transformación: A cada atributo se le resta el mínimo y se divide por el máximo, manteniendo el orden y escalando los resultados por un factor de proporcionalidad.

La plataforma del lado de la demanda provee datos actualizados de las campañas publicitarias en distintos niveles de agregación, en diferentes métricas y diferentes dimensiones de explicación. Para la primera etapa de la elaboración del proyecto se utilizarán los datos agregados a nivel mensual, los cuales tienen el rendimiento de las campañas durante los últimos 3 meses correspondientes a los meses de Julio, Agosto y Septiembre debido a la disponibilidad de datos.

3.1.1. Descripción de los Datos

Los datos proveídos a nivel mensual por *Mediamath* vienen separados por estrategias y por tamaños de creatividades. La separación de los datos y los atributos se muestran a continuación:

3.1.1.1 Dimensiones

A continuación se explican las dimensiones de los datos que provee el DSP *Mediamath*:

- ID_Campaña: Identificador de la campaña publicitaria

- ID_Cliente: Identificador del cliente (como cliente se entiende como la empresa que desea realizar publicidad en internet).
- ID_Estrategia: Identificador de la estrategia de las campañas publicitarias.
- Tamaño_Creatividad: Indica las dimensiones de las creatividades que se utilizaron en las campañas publicitarias.

Estas dimensiones caracterizan la forma de una campaña publicitaria. El acuerdo del nivel de la campaña publicitaria es establecido a nivel global, pero cada campaña tiene distintas estrategias. Cada estrategia tiene sus características únicas y un objetivo único, con el fin de contribuir al rendimiento global de la campaña. Por ejemplo una campaña publicitaria puede tener dos estrategias, en donde la primera se intente lograr la mayor cantidad de impresiones posibles a un bajo costo, mientras que la segunda estrategia sería aparecer en sitios más premium a un mayor costo.

3.1.1.2 Atributos

A continuación se muestran los atributos o características de las campañas publicitarias en internet a través del DSP *Mediamath*:

- Presupuesto_Campaña: Dinero total asociado a una campaña publicitaria (en dólares americanos).
- KPI_Campaña_Tipo: El objetivo de la campaña publicitaria, esta puede ser CPM (costo por mil impresiones), CPC (costo por clic) o CPA (costo por acción).
- KPI_Campaña_Objetivo: El valor numérico del tipo del objetivo de la campaña.
- Presupuesto_Estrategia: Dinero asociado a una de las estrategias de la campaña. La suma de los presupuestos de estrategias no puede superar al de la campaña.
- Fecha_Inicio_Estrategia: Momento en que la estrategia comienza a participar en el protocolo RTB.
- Fecha_Fin_Estrategia: Momento en que la estrategia comienza a participar en el protocolo RTB.
- KPI_Estrategia_Objetivo: El objetivo de la estrategia publicitaria, esta puede ser CPM (costo por mil impresiones), CPC (costo por clic) o CPA (costo por acción).
- KPI_Estrategia_Tipo: El objetivo de la estrategia publicitaria, esta puede ser CPM (costo por mil impresiones), CPC (costo por clic) o CPA (costo por acción).
- Impresiones: valor numérico de las impresiones a realizar.
- Clics: Valor numérico de los clics realizados.

- Gasto_total: Valor numérico del gasto total realizado (en dólares americanos).
- Gasto_total_CPM: Valor numérico del gasto por mil impresiones realizado (en dólares americanos).
- Gasto_total_CPC: Valor numérico del gasto por cada clic realizado (en dólares americanos).
- CTR: Cuociente entre el número de clics y el número de impresiones.
- Gasto_Contextual: Valor numérico del dinero gastado en segmentar los sitios que proveen espacios publicitarios para la venta bajo el protocolo RTB.

En base a estos atributos se generan los siguientes para explicar las campañas publicitarias:

- Duración: Duración temporal de la campaña publicitaria.
- Velocidad de Gasto: Cantidad de unidades monetarias por hora.
- Velocidad de impresiones: Cantidad de impresiones por hora.

Junto con los atributos, se consideró evaluar el tamaño de las creatividades como un factor relevante en las campañas publicitarias. Cada campaña publicitaria tiene un conjunto de creatividades en los cuales se pueden variar en las diferentes dimensiones (ver 2.3.4).

La agregación mensual cuenta con 400 estrategias con variadas configuraciones en la realización de las campañas publicitarias. Las estrategias toman distintos valores para los atributos cuantitativos aún cuando corresponden a los mismos tipos de estrategias. Corresponden a campañas de display⁴.

3.2 Generación de Atributos

Para obtener el resultado intrínseco de cada estrategia es necesario agrupar por el identificador de la estrategia, agrupando el número de impresiones, clics y costos incurridos, debido a que el objetivo de estudio son las estrategias y no las creatividades. Junto con esto, una variable interesante de estudio son el uso o no de distintos tamaños de creatividades. Agrupando las estrategias a nivel mensual y desglosando por las creatividades, obtenemos una variable categórica por si la estrategia utiliza o no cierto tipo de creatividad. El número de creatividades dentro de una estrategia toma valores entre 2-23 diferentes, creando muchas variables nominales que al ser ingresados a un modelo pueden crear problemas de interpretación y atribución. Una forma de reducir el

⁴ Son campañas en donde la creatividad es una imagen o una secuencia dentro del sitio, se excluyen otro tipo de campaña como las de video o audio.

número de variables dicotómicas es agrupándolas, bajo algún criterio para dejar de utilizar cada creatividad como insumo al modelo y utilizar las agrupaciones en vez. Esta información será relevante en los estudios siguientes. El criterio de agrupar los diferentes grupos de creatividades fue la segmentación basada en K-Means y justificada con el criterio Davies Bouldin que se muestran a continuación en la siguiente tabla:

# Cluster	Davies Bouldin
1	0,814
2	0,787
3	0,719
4	0,642
5	∞
6	∞
7	∞
8	∞
9	∞

Tabla 8, Agrupación de Creatividades Según Criterio Davies Bouldin

Por lo que se decide utilizar la segmentación con 5 y el tipo de elementos de los segmentos se muestran en la próxima tabla (para ver el resultado de la segmentación véase anexo [1]) :

Creatividades	Segmento 1	Segmento 2	Segmento 3	Segmento 4	Segmento 5
Tamaños	Pequeñas	Pequeñas	Pequeñas	Grandes	Grandes
Variedad	Media	Alta	Baja	Baja	Alta

Tabla 9, Representación Visual de Segmentos de Creatividades

En donde los dos elementos que distinguen de mejor manera los segmentos son los tamaños de las creatividades y la variedad de los mismos. El tamaño cuantitativo en la tabla representa si la campaña publicitaria contiene creatividades que abarcan una mayor superficie en la pantalla del computador relativo a la creatividad promedio que corresponde a la que tiene 300 x 300 pixeles. Por otro lado la variedad corresponde al número de creatividades que contiene la campaña, lo que muestra que hay campañas con baja variedad y corresponden a aquellas que tienen menos de 5 creatividades.

Por otro lado, la próxima tabla se muestran el número de elementos en cada uno de los segmentos:

# Clusters	Segmentos
5	Segmento1: 101
	Segmento 2: 64
	Segmento 3: 18
	Segmento 4: 67
	Segmento 5: 45

Tabla 10, Cantidad de Elementos en los Segmentos de Creatividades

Luego se agrega como variable nominal el segmento de las categorías de creatividades. El objetivo del proceso anterior es reducir el número de variables categóricas de 23 a 5, con el fin de hacer el resultado más interpretable.

3.3 Validación de Atributos

Para validar cuales de los atributos explican de mejor manera el comportamiento de las campañas publicitarias en internet se desarrolló un análisis de componentes principales, con el objetivo de justificar la utilización de atributos y eliminar aquellos que no necesariamente explican bien el comportamiento. Para que los diferentes órdenes de magnitud no condicionen el resultado del análisis, se realizó una normalización de todos los atributos para que se acoten en el intervalo 0-1. Esta transformación se realizó como lo indica la siguiente fórmula:

$$\bar{X}_i = \frac{X_i - \text{Min}_i(X_i)}{\text{Max}_i(X_i) - \text{Min}_i(X_i)}$$

En donde a cada atributo le resta el mínimo de sus pares, y luego se divide con la diferencia entre el mínimo y el máximo, garantizando valores positivos con un máximo igual a uno y un mínimo con valor cero.

Con ello se espera validar y reducir la dimensionalidad del problema. La siguiente ilustración muestra la proporción acumulada de la varianza explicada por los diferentes vectores propios de la matriz de varianza covarianza:

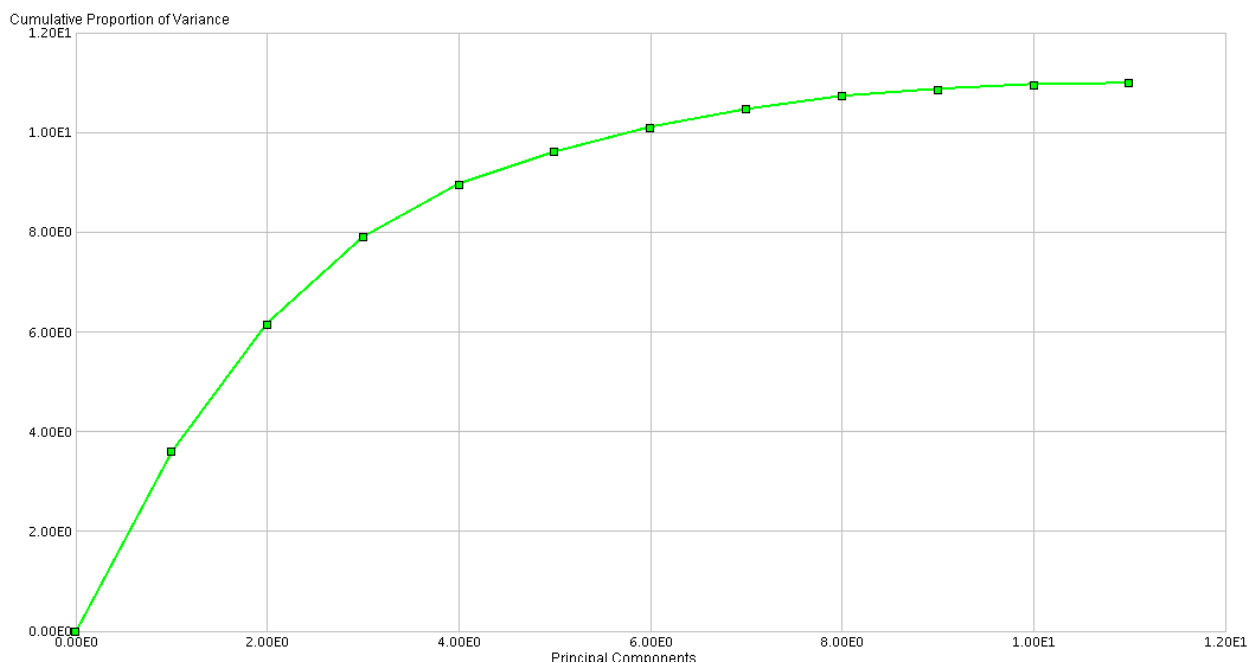


Ilustración 13, Análisis de componentes principales

Este gráfico representa por cada vector propio, que porcentaje de la varianza de los datos se explica con él. Un método tradicional para reducir la dimensión de los atributos y el método para realizar esta operación es elegir los vectores que explican suficiente varianza hasta algún criterio por definir. En este caso no se identifica un salto abrupto de porcentaje de varianza explicada por los vectores, por lo que no es claro con cuales vectores se deben mantener. Luego es necesario identificar los vectores propios y examinarlos detenidamente para evaluar si hay algún atributo poco significativo. En la siguiente tabla se muestran las componentes de los primeros seis vectores del análisis de componentes principales:

Atributo / Vector	PC1	PC2	PC3	PC4	PC5	PC6
strategy_budget	0.332	0.198	-0.093	0.382	-0.490	0.446
strategy_goal_value	-0.183	0.533	0.092	-0.003	-0.149	-0.124
impressions	0.432	0.000	-0.166	-0.156	0.360	-0.250
clicks	0.293	0.313	-0.271	-0.159	0.502	0.410
total_spend	0.439	0.219	-0.137	-0.071	-0.179	0.100
total_spend_cpm	-0.206	0.529	-0.013	0.042	-0.140	-0.146
ctr	-0.181	0.482	-0.029	-0.175	0.192	-0.183
contextual_cost	0.393	-0.024	-0.257	-0.106	-0.347	-0.642
Duracion	-0.033	0.035	-0.250	0.854	0.336	-0.250

Velocidad_Impressions	0.283	0.081	0.604	0.108	0.178	-0.109
Velocidad_Spent	0.284	0.097	0.608	0.122	0.051	-0.053
strategy_budget	0.332	0.198	-0.093	0.382	-0.490	0.446
strategy_goal_value	-0.183	0.533	0.092	-0.003	-0.149	-0.124
impressions	0.432	0.000	-0.166	-0.156	0.360	-0.250
clicks	0.293	0.313	-0.271	-0.159	0.502	0.410
total_spend	0.439	0.219	-0.137	-0.071	-0.179	0.100

Tabla 11, Resultado Análisis de Componentes Principales

Se puede observar que no existe un único atributo despreciable dentro de estos vectores (valores muy cercanos a cero en todos los vectores propios), luego se opta por no eliminar ninguno de los atributos debido a que el método no muestra evidencia de que alguno sea no significativo.

3.3 Segmentación de Estrategias Publicitarias

En base a la experiencia y a la observación de los datos, se muestra evidencia de que hay diferentes comportamientos dentro de las campañas publicitarias dadas sus características, es por ello que se opta por segmentar las campañas publicitarias en base a un comportamiento común.

Para la segmentación de campañas se elabora un método no supervisado de división. Esta clase de métodos no toma la experiencia del negocio y se obtiene intrínsecamente de los datos a partir de un modelo matemático que se explicará a continuación.

Utilizando el mismo criterio para la segmentación de creatividades, se estableció la siguiente tabla comparativa con respecto al número de clusters y a la calidad de ellos en base al indicador Davies Bouldin:

Número de Clusters	Davies Bouldin
2	1,319
3	1,034
4	1,109
5	1.181
6	1.207
7	1.234
8	1.114

Tabla 12, Agrupación de Creatividades Según Criterio Davies Bouldin

Según este análisis, se determina que el número óptimo de segmentos para las estrategias publicitarias son 3 bajo el criterio de Davies Bouldin el cual establece una proporción de la distancia promedio de los datos con respecto a un centroide asociado, dividido por la distancia entre los mismos centroides. Esta métrica busca elaborar segmentos compactos y distanciados entre sí.

Al segmentar las estrategias se obtiene la siguiente tabla donde se muestran los centros de cada clase, con el fin de entenderlos y darles una interpretación intuitiva (los valores son proporcionales debido a que se busca que estén en la misma escala de comparación, por lo que se normalizaron los datos para ajustarlos al intervalo 0-1)-

	Segmento 1	Segmento 2	Segmento 3
strategy_budget	0.659	2.237	-0.302
strategy_goal_value	-0.511	0.746	0.162
impressions	1.145	1.066	-0.443
clicks	0.602	2.360	-0.285
total_spend	0.966	2.818	-0.430
total_spend_cpm	-0.452	0.409	0.151
ctr	-0.334	0.115	0.11
contextual_cost	1.276	-0.015	-0.459
Duracion	0.050	-0.356	-0.007
Velocidad_Impressions	0.043	5.078	-0.163
Velocidad_Spent	0.043	5.372	-0.171

Tabla 13, Resultado Segmentación Estrategias Publicitarias

Con estos resultados, se pueden separar los segmentos por sus cualidades más diferenciadoras entre ellas para entregar características intuitivas de cada una. El desglose se muestra en la siguiente tabla:

Segmento 1	Segmento 2	Segmento 3
Publicidad Constante	Promoción Premium	Experimentación
Presupuesto Medio	Presupuesto Alto	Presupuesto bajo
Impresiones Costo Bajo	Impresiones Costo Alto	Impresiones Costo Medio
Larga Duración (Mes Completo)	Corta Duración (1-5 días)	Media Duración (5-20 días)

El segmento 1 se interpreta como campañas de presupuesto promedio entre las demás, que busca impactar a la mayor cantidad de usuarios a un bajo costo, por lo que no espera comprar inventario considerado de mayor calidad por que el mismo mercado impone un precio mayor a estas oportunidades de compra. Son campañas de larga duración temporal y tiene un ritmo de gasto promedio bajo. Un ejemplo de este tipo de campañas son productos conocidos que son constantemente publicitados para recordar a los usuarios finales. Como ejemplo son las campañas de Dettol utilizadas como ejemplo previamente en el informe, las cuales buscan impactar a la mayor cantidad de usuarios a el menor costo posible.

El segundo segmento corresponde a campañas publicitarias con un alto presupuesto y un alto nivel de inversión en la compra de espacios publicitarios. Pretende llegar a inventario considerado por el mercado como Premium en poco tiempo y a una alta velocidad. Estas son campañas esporádicas de eventos particulares de una marca en la cual necesitan hacer publicidad, como por ejemplo las campañas realizadas por el

avisador Ideal para la campaña publicitaria del Energy Race de Agosto, con la que desean buscar suscriptores rápidamente para inscribirse en la carrera.

Finalmente el tercer segmento son campañas de un presupuesto bajo, a un nivel de impresiones promedio bajo y tienen una duración temporal promedio. Estas campañas se pueden interpretar de dos maneras diferentes. Son realizadas por clientes con poca experiencia en la compra programática, por la que destinan poco o bajo presupuesto a intentar realizar la compra y esperan a ver los resultados. Esto ocurre por ejemplo en el lanzamiento de un nuevo sitio de la Cámara Chilena de la Construcción a la cual se le ofreció un plan de medios incluyendo compra programática y fue aceptado. También se interpreta como campañas publicitarias con un segmento de clientes objetivo muy pequeño, con lo cual solo desea impactar en sitios en los que pocos usuarios ingresan pero son su segmento objetivo, para ilustrar esta la empresa Kaufmann la cual vende camiones de alto costo y su segmento de impresión son sitios de camiones los cuales son reducidos en número de sitios y en tráfico de usuarios en los sitios.

3.4 Supuesto Cualitativo

Es natural pensar que tanto el contenido como la calidad del mensaje enviado de las creatividades tenga un fuerte impacto en el comportamiento de la campaña publicitaria. Es decir que una mejor publicidad o un mejor mensaje afecta directamente sobre las métricas de rendimiento a lo largo de la ejecución de la campaña. Este atributo cualitativo no será integrado en el modelo predictivo explicado por el proceso realizado por la empresa en la negociación de valores de atributos, y por la carencia del valor del atributo explícito.

En la empresa *Affiperf* no realiza un análisis previo sobre el mensaje forma y contenido de la creatividad al realizar sus proyecciones de la campaña publicitaria. Se guía en base a datos históricos, a la segmentación y presupuesto del cliente para el nivel de impresiones. Para el CTR (número de clics por volumen de impresiones) lo mantienen fijo en 0,11% o utilizan información histórica sobre las campañas que tienen información histórica, es decir que utilizan las métricas que se obtuvieron en la realización de la misma campaña. Por otro lado el CPC (costo por clic) se elabora como una combinación del costo por mil impresiones (CPM) y el ratio de clics establecido. Es por esto que no se agrega al modelo alguna métrica de calidad, debido a que no existe una métrica establecida para cuantificar el nivel del mensaje y no es el foco de estudio realizar este análisis.

Considerando que el atributo de la calidad del mensaje enviado puede ser bastante representativo de la campaña publicitaria, no agregarlo al modelo tiene la única implicancia de no capturar el nivel de explicación en esa dimensión y no condiciona los resultados de los otros atributos en una regresión lineal [12], debido a que no existe una correlación evidente entre los atributos explicados para el modelo con la calidad del mensaje en términos de las variables que controla *Affiperf* debido a que no lo considera actualmente en el proceso. Por otro lado, el presupuesto y la duración podrían estar

correlacionados con el mensaje debido a que la inversión en crear el mensaje podría estar correlacionado directamente. Esta dimensión no se aplica al presente informe y podría ser integrado en un nivel futuro. Si hubiera algún tipo de correlación, el resultado de la regresión tendría un sesgo debido a que no se cumple con la independencia del error en la regresión lineal con el resto de los atributos. Es posible que se dé este caso en situaciones en cuanto la calidad del mensaje sea lo bastante buena como para cambiar la configuración de la campaña publicitaria de función del contenido de la creatividad, pero no es un caso común según la experiencia del negocio y del tutor de la memoria.

Capítulo 4. Valoración Económica de la Predicción

Cada implementación de campaña publicitaria tiene por detrás una negociación realizada entre *Affiperf* y la empresa que desea comunicar su mensaje vía la compra programática. Esta negociación condiciona la forma de elaborar la campaña publicitaria según los intereses de cada ente participante. Para ilustrar los términos de la negociación, la segunda fila de la siguiente tabla muestra una versión simplificada de los términos de negociación que se establecen previo a una campaña publicitaria (véase también ilustración 9 con un ejemplo concreto):

	Nivel de Impresiones	Clics	CTR	CPC	CPA	Inversión
Proyectado						
Obtenido						

Tabla 14, Esquema de Proyección de la Negociación de las Campañas Publicitarias

La tercera fila de la tabla se elabora una vez realizada la campaña publicitaria. Luego se considera que el avisador queda conforme si se logran todas las métricas proyectadas al final de la campaña. Luego para poder valorizar el riesgo de realizar malas proyecciones, es necesario evaluar al cliente y evaluar la probabilidad de que el cliente deje de serlo debido a proyecciones poco representativas.

4.1 Valorización de los Clientes

Para valorizar los clientes se realizó una segmentación por niveles de inversión, para ponderar el valor mensual de cada tipo de cliente de manera separada. Junto con esto se calcula la frecuencia en la cual se presenta cada segmento para finalmente calcular un valor esperado de cada cliente mensualmente.

En el año 2015 hasta Octubre *Affiperf* ha trabajado con 46 clientes diferentes, con más de 100 campañas publicitarias mensuales de las cuales el 58% de ellas corresponden a campañas de display las cuales son el objeto de estudio del proyecto. En la siguiente ilustración se muestra el distinto nivel de utilidad que generan las empresas avisadoras a *Affiperf* :

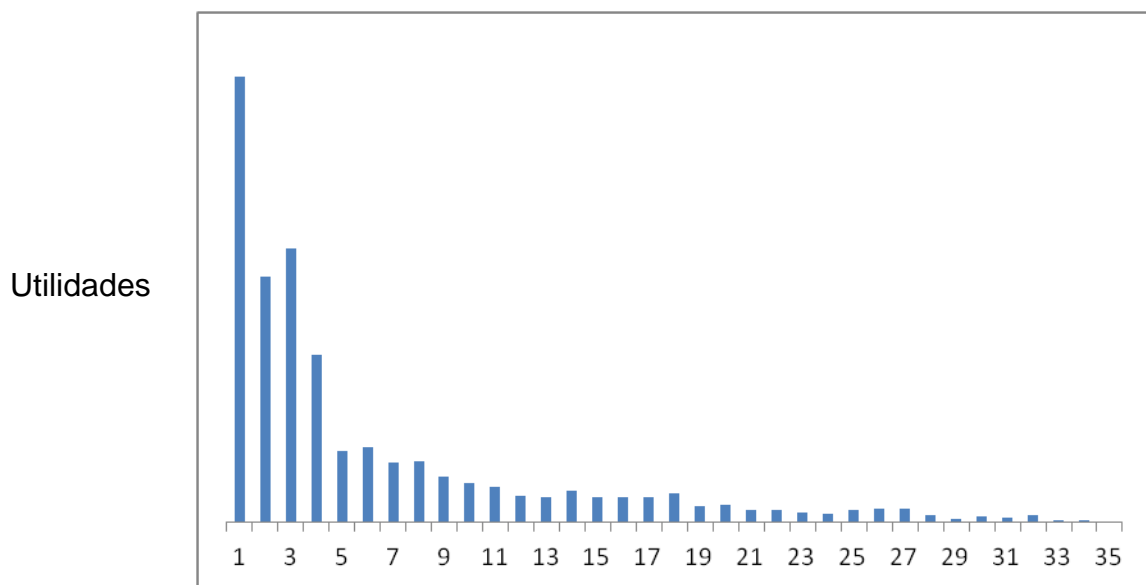


Ilustración 14, Distinta Utilidades Según Clientes

En el cual el eje x corresponde a un número identificador de la empresa y el eje y representa las utilidades dentro del periodo del primero de Julio hasta el 31 de Septiembre del año 2015. Los valores son confidenciales pero el objetivo de la ilustración es mostrar que hay distintos grupos de clientes en función del retorno que genera para *Affiperf*. Los primeros 4 clientes generan un nivel de utilidades cercano a 3 veces más que los siguientes 12. Los últimos generan un bajo nivel marginal en relación a los demás. Dado esto, cabe la necesidad de incluir a la valorización esta división desigual de los ingresos generados por los clientes.

Luego separamos los clientes en tres clases, asociando el valor promedio de cada uno de los grupos y el porcentaje total de la inversión que representan. La siguiente tabla ilustra los distintos tipos de clientes de la empresa de compra programática a nivel mensual:

	% de la Inversión	% de los clientes	Utilidad Promedio [UM]
Segmento 1	60,92%	11,43%	4.609
Segmento 2	32,88%	45,71%	1.104
Segmento 3	6,20%	42,86%	236

Como refleja la tabla y el gráfico, la mayor parte de la inversión está concentrada en poco clientes, mientras que la mayor cantidad de clientes (que son el segmento 2 y 3) unidos constituyen cerca del 90% de los clientes pero no llegan a ser el 40% de la inversión.

Considerando las grandes diferencias en las empresas que trabajan con compra programática, para tomar un número representativo de cuánto vale un cliente en términos monetarios se pueden utilizar una gran variedad de técnicas. Entre ellas se utilizará en primera instancia el valor esperado de un cliente en términos de las utilidades generadas para *Affiperf* según las proporciones (el cual es la multiplicación por fila y suma por columna de las últimas dos columnas de la tabla anterior) lo cual entrega un valor esperado de los clientes de 1.132 [UM] mensualmente. Luego para evaluar el valor de la vida de los clientes se considerará como ciclo de vida del cliente (son los años que actualmente lleva operando la empresa) según la siguiente fórmula de valor presente[13]:

$$VP = \sum_{t=1}^t \frac{vm}{(1+r)^{t-1}}$$

En donde VP representa el valor presente del cliente, *vm* es el valor mensual del cliente y *r* representa la tasa de descuento por el valor futuro del dinero. Esta fórmula captura que el valor del dinero es menos mientras el flujo se lleve a cabo más tarde en el tiempo considerando a 3 años de vida del cliente.

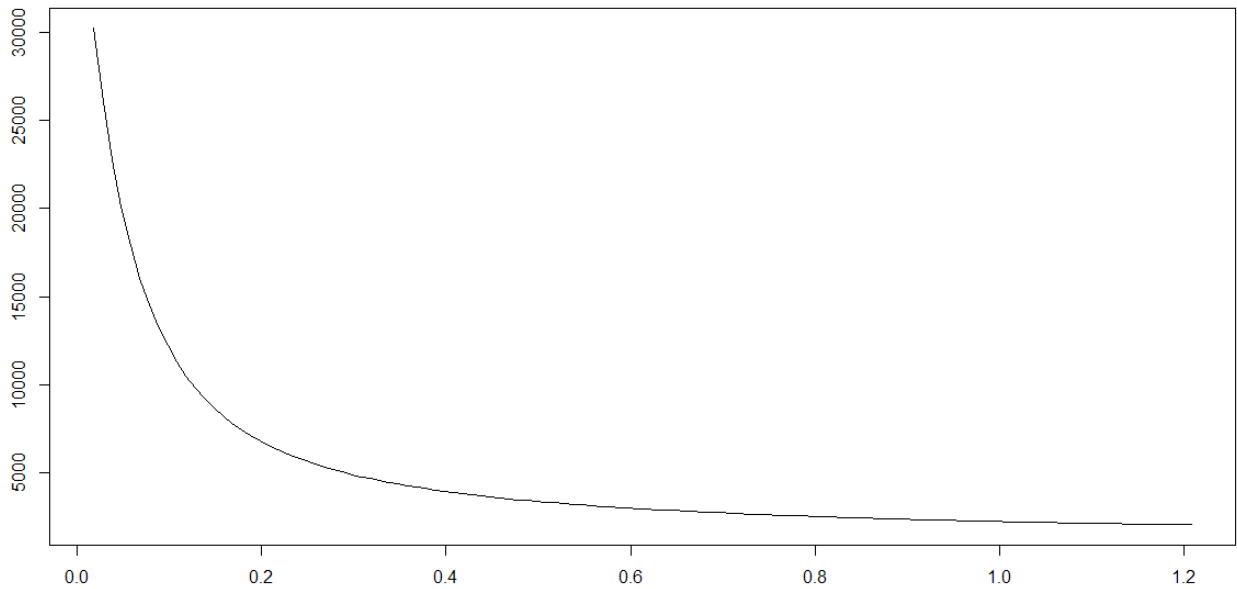


Ilustración 15, Valor Presenta a Diferentes Tasas Mensual

La ilustración 15 muestra la elasticidad del valor presente de un cliente promedio variando la tasa de descuento con la que se transforman los flujos de dineros del futuro al día de hoy. Notamos que es decreciente y tiene una pendiente más pronunciada en los primeros valores de la tasa de descuento (es de esperarse debido a que nace de una suma geométrica). La variabilidad en los primeros valores muestra que el análisis puede cambiar significativamente según la apreciación de los valores futuros del dinero. Se elige una tasa de descuento del 10% compuesta anual (0,79% mensual), por ser un valor referencial clásico en el uso de tasas de descuento.

Por otro lado es necesario considerar el umbral temporal de la vida de los clientes con la empres *Affiperf*. Es un valor difícil de evaluar debido a que la empresa lleva únicamente 3 años funcionando en Chile. La siguiente ilustración muestra la sensibilidad del valor del cliente con diferentes números de periodos (anuales) con una tasa de descuento del 10%:

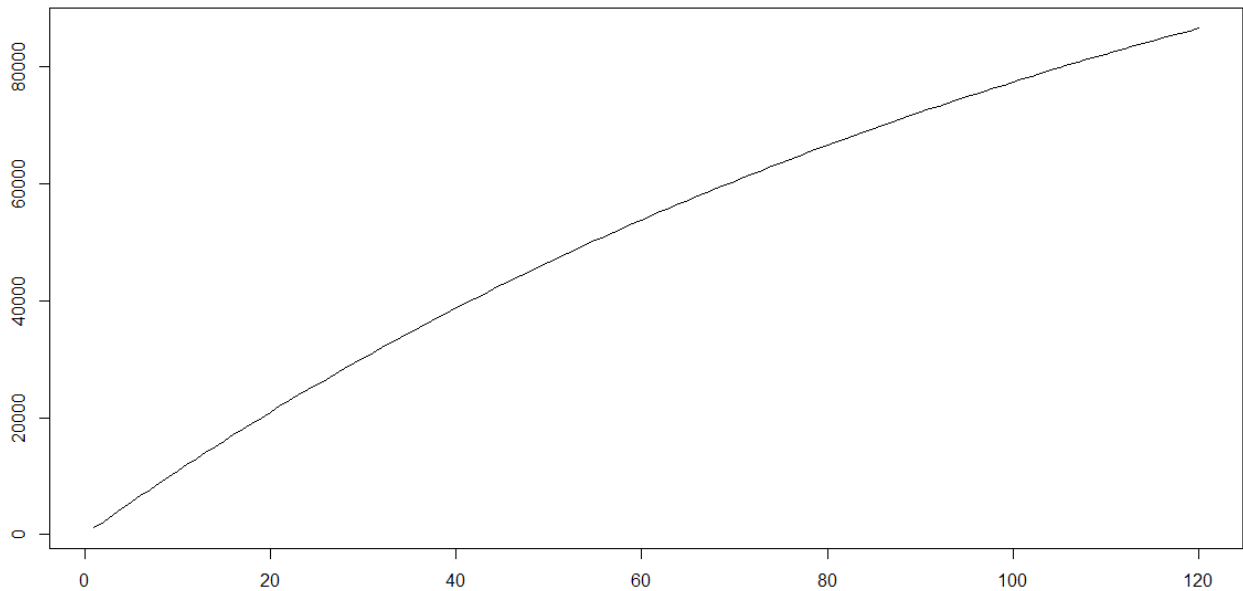


Ilustración 16, Valor Presente, Distintos Periodos Mensuales

Luego se puede apreciar que el valor presente tiene una forma cóncava y es muy volátil en los primeros periodos a considerar. Por lo mismo el análisis expuesto en los modelos realizados en el informe pueden variar drásticamente dependiendo del número de periodos (mensuales) que se consideren. El ciclo de duración de un cliente en *Affiperf* varía considerablemente, pues depende de los planes publicitarios que tenga la empresa avisadora. Esto valida la hipótesis de que los clientes tienen un comportamiento errático se considerarán 3 años (lo que equivale a 36 periodos mensuales).

Finalmente el valor de un cliente se evaluó con un periodo de 3 años con una tasa de descuento del 10% y se obtuvo un valor de 35.393 [UM].

4.2 Probabilidad de Deserción de un Cliente

Ante la situación en la que a un cliente no se le cumplan las proyecciones de una campaña mensual, existen dos posibles escenarios. El primero es que el cliente continúe usando el servicio de compra programática es sus planes de publicidad y la segunda es que deje de utilizarlos en función de lo que él considera un pacto deficiente ante el no cumplimiento de las proyecciones de las campañas publicitarias. Es por esto que para poder cuantificar el costo de que a un cliente se le generen proyecciones y no se cumplan, se debe evaluar la probabilidad de que este abandone el servicio completamente.

Es claro que esta decisión no ocurre únicamente por la falta a la proyecciones acordadas, sino que pueden ocurrir diversos factores que afecten a la decisión de

utilizar o no el servicio prestado por la empresa *Affiperf*, pero para fines del estudio se considerará la información histórica de deserción.

Como deserción se considerará como una empresa que deje de realizar compra programática por dos meses luego de haber sido expuesto a una proyección no alcanzada y deje de utilizar a la empresa en los siguientes dos periodos. La siguiente ilustración muestra el número de campañas publicitarias por cliente:

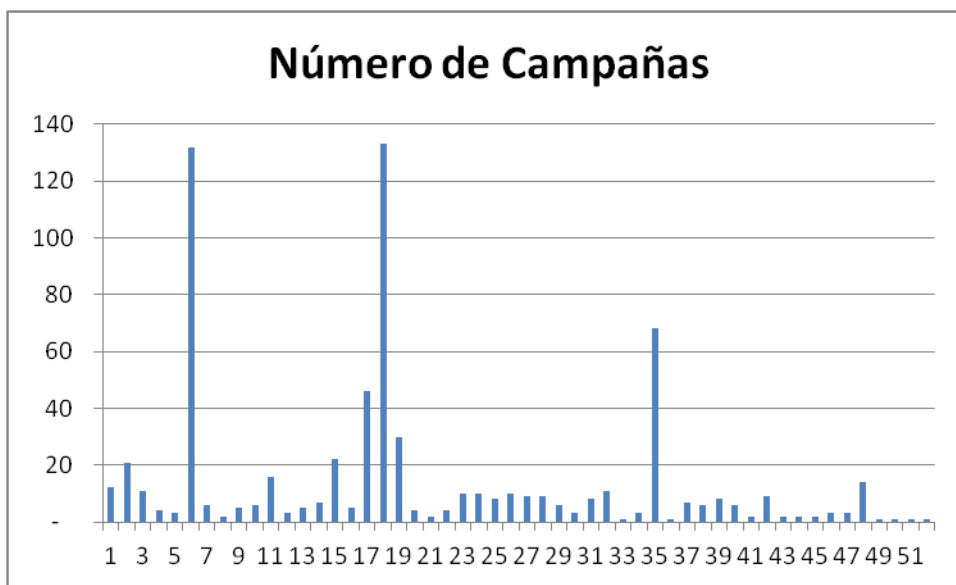


Ilustración 17, Número de Campañas

El gráfico ilustra el número de campañas publicitarias por cada cliente. Hay una natural varianza entre ellas explicadas por la diferencia en la inversión de las empresas (véase 4.4.1 valorización de los clientes). Excluyendo los avisadores que retoman el uso de la empresa para campañas publicitarias, existe un porcentaje de deserción del 1,25% (en la práctica fueron 6 clientes correspondientes al 11,1% del total, sobre la cantidad de oportunidades de deserción las cuales corresponden a 477) a lo largo del año 2015 desde el primero de enero hasta el 31 de septiembre. Este número es calculado como la suma de los clientes que cumplen con el criterio de deserción dividido por el total de clientes activos en el periodo en cada periodo.

4.2 Costos de la Predicción

Realizados los pasos anteriores, se puede generar una tabla de costos y beneficios de realizar campañas publicitarias si se acierta en la predicción de los rendimientos futuros de las campañas o no. La siguiente tabla muestra los diferentes escenarios a los que se puede enfrentar un cliente de la compra programática luego de haber concluido el

periodo de la campaña publicitaria en cuestión, considerando los escenarios favorables y los que no lo son.

Predicción	Beneficio/Costo
Sub-estimación	<ul style="list-style-type: none"> • Posibles utilidades • Posibles pérdidas de operación
Sobre-estimación	<ul style="list-style-type: none"> • <i>Posible pérdida del cliente</i> • Posibles utilidades • Posibles pérdidas de operación

En el caso en que *Affiperf* subestima los resultados de sus campañas en términos de CPC y CTR obtiene mejores métricas para la realización de la campaña por lo que se considerará que el cliente queda satisfecho con el acuerdo logrado. En este caso existe la posibilidad de marginar obteniendo el resultado de la campaña a un menos costo del que se compra, luego *Affiperf* obtiene un utilidades directas del cliente. Para calcular el costo de subestimar el resultado de las campañas se calcula como la cantidad de inversión en la campaña publicitaria extra gastado, luego de completar la meta de la campaña. En el segundo caso en el que obtuvo el resultado de la campaña a un precio mayor, la empresa puede generar pérdidas. Esto ocurre cerca del 2% de los casos en la base de datos del 2015.

Por otro lado si sobre estima el resultado de la campaña, el cliente queda insatisfecho con el resultado final y por lo tanto se puede perder al cliente. El valor monetario de la pérdida del cliente se puede calcular como la probabilidad de que el cliente se vaya luego de un incumplimiento del acuerdo con *Affiperf* el cual fue calculado en la sección 4.2 multiplicado por el valor de la vida del cliente calculado en 4.1 obteniendo un costo esperado de 442,42 [UM] además de obtener o no margen en la campaña publicitaria.

Luego la función de costos queda de la siguiente manera:

$$F(x) = \begin{cases} 442,42 & x < \bar{x} \\ c * (x - \bar{x}) & x > \bar{x} \end{cases}$$

en donde \bar{x} es el resultado de la métrica de rendimiento se la campaña publicitaria real, x es la predicción de la campaña y c es el costo por exceder la campaña publicitaria.

A modo de ejemplo considere la siguiente proyección para una campaña publicitaria:

	Nivel de Impresiones	CTR	CPC	Clics	Inversion Cliente	Inversión <i>Affiperf</i>
Proyectado	100.000	0,0011	9,09	110	1.000	
Obtenido	100.000	0,002	5	200		800

Tabla 15, Esquema de Proyección de la Negociación de las Campañas Publicitarias

Si la campaña tiene como objetivo primario clics, *Affiperf* incurre en un sobre costo de 90 clics por sobre lo proyectado, reduciendo su margen puesto a que pudo haber logrado el nivel de clics con 550.000 impresiones.

Capítulo 5. Evaluación de Modelos Predictivos

La evaluación cuantitativa de la metodología se desarrolla a partir de la aplicación de la misma y compararla con los resultados actuales de la empresa. La manera actual se basa en la opinión de expertos y se pretende comparar con los resultados de los modelos aplicados según lo descrito en el informe.

Los modelos elegidos son regresiones lineales a los atributos considerados en el pre procesamiento de los datos. Para ello se elaborará una descripción del valor agregado del modelo según sus características y la valorización económica previamente realizada. Para entrenar los datos se utilizaron los datos desde el primero de julio hasta el 31 de septiembre del año 2015, y como set de evaluación se aplicó el modelo para predecir en los datos de agosto. La siguiente ilustración muestra la separación de los datos:

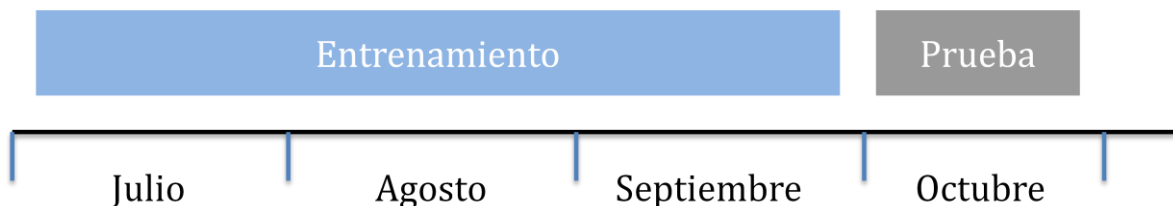


Ilustración 18, Separación datos Entrenamiento y Prueba

5.1 Predicción de CTR

5.1.1 Resultados sin Segmentación

En primera instancia se evaluará el modelo con los atributos descritos a todo el universo de estrategias publicitarias, para esto se utilizará como variable dependiente el CTR y como variables independientes las obtenidas en el capítulo 3.

Al aplicar la regresión lineal sobre los datos de entrenamiento se obtienen los coeficientes asociados a los atributos mencionados tomando como variable objetivo el CTR [ver anexo 2].

Esto estimados se obtienen en base al método de mínimos cuadrados ordinarios (R-cuadrado de 0,49, luego los atributos capturan el 49% de la varianza de los datos). A continuación se analizará cada uno de los resultantes valores asociados a cada atributo:

- *total_spend_cpm*: El valor positivo recalca que a mayor gasto por la impresión, se obtiene un mejor rendimiento en términos de CTR. Probablemente debido a que los espacios publicitarios más costosos tienen mejor visibilidad ante el usuario final.
- *contextual_cost*: El valor positivo indica que mientras más se invierta en segmentar los sitios según bases de datos de proveedores, se mejora la variable en cuestión. Esto aunque la magnitud del estimador asociado es bajo en comparación con los otros.
- *Velocidad_Impressions*: A mayor velocidad de impresiones, mejor será el resultado de la campaña. Esto se explica a que los usuarios podrían ser impactados en poco tiempo varias veces por la publicidad y por lo tanto reaccionar ante ella.
- *Velocidad_Spent*: A mayor gasto de dinero se obtiene un menor rendimiento. Es un resultado de una variable no significativa.
- *total_spend*: Otro parámetro poco significativo, este puede explicarse de que las personas pueden ser independientes del número de apariciones y la visibilidad es capturada por el primer parámetro y no por este.
- *Duracion*: No ejerce mayor efecto en la variable explicativa.
- *impressions*: No ejerce mayor efecto en la variable explicativa.

A continuación se realizará una validación económica de la predicción del modelo con respecto a lo obtenido realmente en el mes de octubre.

5.1.1.1 Costo de Sobre Estimación de las de Proyecciones de Métricas de las Campañas Publicitarias CTR

Para evaluar el costo de subestimación de las campañas publicitarias, se compara la cantidad de veces que el modelo subestima las campañas en relación a lo realizado por *Affiperf*.

Sobre estimados por <i>Affiperf</i>	50,6%
Sobre estimados por el modelo	46,8%

Tabla 15, Sobreestimación de las Métricas de Rendimiento de las Campañas Publicitarias Reales vs Modelo

Lo que se traduce en costos evaluado por el costo esperado de la pérdida de un cliente como:

Pérdidas estimadas <i>Affiperf</i>	223,86 [UM]
Pérdidas estimadas el modelo	207.58 [UM]

Tabla 16 Sobre Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo

Como se puede apreciar de este modelo sin segmentar por tipo de segmento, el ahorro potencial es de 16,28 [UM].

5.1.1.2 Costo de Sub Estimación de Proyecciones de las Métricas de las Campañas Publicitarias CTR

El cálculo de sub estimar las métricas de rendimiento de las campañas es el costo de sobre invertir en campañas que ya realizaron sus proyecciones, por lo que es un margen el cual *Affiperf* pierde. La siguiente tabla muestra la comparación entre la pérdidas reales y las obtenidas por el modelo.

Pérdidas estimadas <i>Affiperf</i>	334,87 [UM]
Pérdidas estimadas el modelo	329,49 [UM]

Tabla 17, Sub Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo

Por lo que obtenemos que el modelo mejora las utilidades de la empresa en 5,38 [UM].

Finalmente el Modelo tiene un menor costo tanto en la sobre y la subestimación de las campañas publicitarias, reduciendo el monto final en 21,66 [UM].

5.1.2 Resultados con Segmentación

A continuación se muestran los resultados luego de realizar la segmentación en las campañas publicitarias. De igual modo que en los resultados sin segmentación, se presentarán los costos de sobre estimar las campañas primero, luego el costo de subestimar las campañas para finalmente incorporar los resultados con el fin de valorar el modelo. Nuevamente la variable dependiente será el CTR y las variables independientes las obtenidas en el capítulo 3.

5.1.2.1 Costo de Sobre Estimación de las de Proyecciones de Métricas de las Campañas Publicitarias CTR

En la evaluación del costo de sobre estimar las métricas de rendimiento de las campañas publicitarias, se aplicó un modelo de regresión dentro de cada uno de los segmentos generados en 3.3 y se valorizó de la misma manera que el punto 5.1.1. La siguiente tabla muestra el resumen de los resultados comparando el modelo con las predicciones de *Affiperf* en el mes de octubre.

Segmento	% Campañas Subestimadas Affiperf	% Campañas Subestimadas Modelo	Costo Affiperf [UM]	Costo Modelo [UM]	Margen Potencial [UM]
Publicidad Constante	35%	47%	156.55	234.23	-78.07
Promoción Premium	59%	48%	261.06	214.35	48.67
Experimentación	54%	32%	239.35	145.06	94.27

Tabla 18, Sobre Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo

En donde se recalca que el modelo obtiene un mayor rendimiento en los segmentos de promoción premium y experimentación. Esto se debe principalmente a que las predicciones de *Affiperf* para campañas que no tienen precedentes es basado en la experiencia global del encargado de negociación, mientras que las campañas que tienen publicidad constante *Affiperf* utiliza información histórica para establecer sus predicciones las cuales superan al modelo debido a que es información específica de la campaña ya realizada con anterioridad.

5.1.2.2 Costo de Sobre Estimación de Proyecciones de las Métricas de las Campañas Publicitarias CTR

Los resultados sobre cada tipo de campaña publicitaria se realiza separadamente en cada uno de los segmentos y se resume en la siguiente tabla:

Segmento	Costo Affiperf [UM]	Costo Modelo [UM]	Margen Potencial [UM]
Publicidad Constante	52,29	37,02	-15,02
Promoción Premium	261,06	214,35	46,71
Experimentación	112,85	90,02	22,83

Nuevamente se obtiene el mismo patrón en el cual *Affiperf* genera mejores proyecciones en publicidad que ha realizado con anterioridad puesto a que utiliza información histórica para la negociación previa al modelo. Esto ocurre en campañas publicitarias que son constantes en el tiempo el cual representa el primer segmento de la tabla. Por otro lado el modelo obtiene una mejor predicción que *Affiperf* en los segmentos de campañas que no se han realizado y por lo tanto utiliza la experiencia global en la negociación.

A modo de resumen, los resultados obtenidos según la segmentación genera la siguiente tabla de márgenes potenciales:

Segmento	Margen Potencial [UM]
Publicidad Constante	-93,09
Promoción Premium	95,43
Experimentación	117,1

Tabla 19, Sub Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo

5.2 Predicción de Costo por Clic (CPC)

Para la métrica del rendimiento del CPC (costo por clic) se seguirá el mismo procedimiento que en la sección 5.1. En un principio se evaluarán los resultados previos a la segmentación de las campañas publicitarias, se evaluarán los resultados y se compararán con los resultados obtenidos actualmente por *Affiperf*.

5.2.1 Resultados sin Segmentación

Se aplica el modelo al conjunto total de las estrategias publicitarias para obtener las elasticidades (coeficientes asociados a las variables explicativas) y se obtiene los resultados [ver anexo 3].

Esto estimados se obtienen en base al método de mínimos cuadrados ordinarios (R-cuadrado de 0,038, luego los atributos capturan el 3,8% de la varianza de los datos. De la misma manera que el apartado anterior, se desglosará el efecto y se entregará una interpretación de la magnitud y signo del mismo.

- total_spend_cpm: Valor positivo pero estadísticamente poco significativo.
- contextual_cost: Valor positivos y poco significativo, luego invertir en utilizar categorización de datos no influye directamente en el costo por clic.

- Velocidad_Impressions: Valor positivo y poco significativo. Luego entregar una alta cantidad de impresiones en un corto periodo de tiempo no se relaciona con la variable objetivo.
- Velocidad_Spent: Valor negativo y poco significativo, por lo que de nuevo no hay relación con la cantidad de dinero invertido en la compra programática con el costo por clic.
- total_spend: Valor positivo y con bajo nivel de significancia, lo que podría significar que aún cuando un usuario sea impactado más veces su decisión de hacer clic o no, es invariante.
- Duracion: No ejerce mayor efecto en la variable explicativa.
- impressions: Valor negativo y un alto nivel de significancia pero con un valor de un orden de magnitud mucho menor. Esto podría explicar porque un usuario no se ve afectado en su decisión a pesar de ser expuesto a la creatividad varias veces.

A continuación se realiza una validación económica de la predicción del modelo con respecto a lo obtenido realmente en el mes de octubre para el cpc.

5.2.1.1 Costo de Sobre Estimación de las de Proyecciones de Métricas de las Campañas Publicitarias CPC

Para evaluar el costo de subestimación de las campañas publicitarias, se compara la cantidad de veces que el modelo subestima las campañas en relación a lo realizado por *Affiperf*.

Sobre estimados por <i>Affiperf</i>	67,3%
Sobre estimados por el modelo	85,3%

Tabla 20, Sobreestimación de las Métricas de Rendimiento de las Campañas Publicitarias Reales vs Modelo

Lo que se traduce en costos evaluado por el costo esperado de la pérdida de un cliente como:

Pérdidas estimadas <i>Affiperf</i>	298,04 [UM]
Pérdidas estimadas el modelo	337,04 [UM]

Tabla 201, Sobre Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo

Como se puede apreciar de este modelo sin segmentar por tipo de segmento, el ahorro potencial es de -39 [UM].

5.2.1.2 Costo de Sub Estimación de Proyecciones de las Métricas de las Campañas Publicitarias CPC

El cálculo de sub estimar las métricas de rendimiento de las campañas es el costo de sobre invertir en campañas que ya realizaron sus proyecciones, por lo que es un

margen el cual *Affiperf* pierde. La siguiente tabla muestra la comparación entre la pérdidas reales y las obtenidas por el modelo.

Pérdidas estimadas <i>Affiperf</i>	754,14 [UM]
Pérdidas estimadas el modelo	706,38 [UM]

Tabla 2221, Sub Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo

Por lo que obtenemos que el modelo mejora las utilidades de la empresa en 5,38 [UM].

Finalmente el Modelo tiene un mayor costo en la sobre estimación y menor en la subestimación de las campañas publicitarias, aumentando del monto final en 8,76 [UM].

5.2.2 Resultados con Segmentación

Nuevamente se desglosan los diferentes costos para los resultados con segmentación para el costo por clic.

5.2.2.1 Costo de Sobre Estimación de las de Proyecciones de Métricas de las Campañas Publicitarias CPC

En la evaluación del costo de sobre estimar las métricas de rendimiento de las campañas publicitarias, se aplicó un modelo de regresión dentro de cada uno de los segmentos generados en 3.3 y se valorizó de la misma manera que el punto 5.1.1. La siguiente tabla muestra el resumen de los resultados comparando el modelo con las predicciones de *Affiperf* en el mes de octubre.

Segmento	% Campañas Subestimadas <i>Affiperf</i>	% Campañas Subestimadas Modelo	Costo <i>Affiperf</i> [UM]	Costo Modelo [UM]	Margen Potencial [UM]
Publicidad Constante	70,5%	76,4%	312,31	338,34	-25,02
Promoción Premium	59%	79,5%	261.06	351.75	-90,69
Experimentación	67,21%	88,02%	297.36	390.23	92.87

Tabla 2322, Sobre Estimación de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo

Se puede observar que el rendimiento del modelo implementado es deficiente con respecto a lo actualmente elaborado por *Affiperf*, en cada uno de los segmentos. Esto se explica porque el costo por clic tiene diversas fuentes de incertidumbre lo que implica que las variables utilizadas en el modelo no logran capturar todas las dimensiones de la varianza de la nube de datos, entregando resultados ineficientes.

5.2.2.2 Costo de Sobre Estimación de Proyecciones de las Métricas de las Campañas Publicitarias CPC

Los resultados sobre cada tipo de campaña publicitaria se realiza separadamente en cada uno de los segmentos y se resume en la siguiente tabla:

Segmento	Costo Affiperf [UM]	Costo Modelo [UM]	Margen Potencial [UM]
Publicidad Constante	153,19	203,31	-50,12
Promoción Premium	261,06	351,63	-127,04
Experimentación	37,72	134,02	-96,3

Se puede ver que como explican todos los resultados, el modelo propuesto establece una peor predicción en cada uno de los segmentos debido a la variabilidad en los datos, es por lo que se recomienda utilizar la información histórica de cada campaña y la experiencia del encargado de negociaciones para lograr mejores retornos para *Affiperf* en la negociación previa a la realización de las campañas publicitarias.

Capítulo 6. Conclusiones

6.1 Conclusiones de la Metodología

El trabajo realizado obtiene como fin último establecer una metodología basada en información histórica para mejorar las negociaciones en las campañas publicitarias realizadas en internet en base al protocolo de subasta en tiempo real a través de la empresa *Affiperf*. La aplicación de la metodología se separa en dos etapas que aglomeran la elaboración del proyecto. La primera etapa radica en valorizar los costos directos en el proceso de negociación de las campañas publicitarias, considerando tanto los costos de subestimar y sobre estimar las proyecciones de las distintas métricas de rendimiento en las campañas publicitarias. La segunda etapa es mejorar la predicción de las distintas campañas publicitarias para proponer a los clientes de la empresa resultados realistas sobre las métricas de rendimiento de las campañas publicitarias para que en la implementación de las campañas logre las proyecciones y los clientes queden conformes con los resultados.

La primera etapa se realizó separado en dos diferentes pasos, correspondientes a los dos distintos costos en los que se incurre. Para evaluar el costo de sobre estimar las métricas de rendimiento se consideró la probabilidad de que un cliente deje de implementar campañas publicitarias en internet vía *Affiperf*, perdiendo de esta manera a un cliente y todo el valor (utilidades) que genera a lo largo del ciclo de vida del cliente mismo. Para calcular esta probabilidad fue necesario recurrir a información histórica de la deserción de los mismos que puedan ser adjudicados por el incumplimiento de las proyecciones de los rendimientos de las campañas publicitarias. Se elaboro un criterio de adjudicación en el cual el cliente deja de realizar campañas publicitarias luego de tener dos campañas que no cumplen con las expectativas en un periodo de dos meses.

En función de esto, se determino que 6 clientes cumplen con esta condición en las 117 oportunidades obteniendo un valor promedio del 1,25% de los casos. Este valor es netamente referencial pero fue validado por el tutor de la memoria. Es posible refinar este número utilizando otros criterios e ir actualizando el valor a medida que avance el número de campañas y la deserción de los clientes. Una vez obtenida la probabilidad, se calcula el valor de la vida del cliente para *Affiperf* el cual se calcula como el promedio ponderado de la utilidad que genera cada cliente mensualmente y se lleva a valor presente con una tasa de descuento del 10% compuesto anual (0,79% mensual) y se calcula el ciclo de vida a lo largo de 3 años. Los tres años se deben a la duración de *Affiperf* en el mercado chileno. Éste número puede ser actualizado en cuanto la participación de la empresa en el mercado tenga mayor información histórica. Se obtiene un valor utilitario para la empresa de 35.393 [UM], llevando a un valor económica de la pérdida de los clientes en cada negociación de 442,44 [UM].

La evaluación del costo de sub estimar las métricas de rendimiento de las campañas publicitarias se elabora como la sobre inversión de *Affiperf* para cumplir y exceder las proyecciones realizadas en la negociación. si se proyecta una métrica de un valor X y se obtiene un valor de $X+\epsilon$ se está excediendo la inversión en los DSP y por lo tanto *Affiperf* incurre en una reducción de su margen y por lo tanto un costo de oportunidad. Para evaluar tal situación se calcula el excedente de inversión por el costo que genera el DSP. Obteniendo cuánto dinero se perdió del margen en una campaña cuyas proyecciones ya han sido alcanzadas.

La segunda etapa del informe es establecer cuáles son las características de cada campaña publicitaria, con las cuales se puede predecir el comportamiento de la mismas. Para esto se tomaron algunos de los atributos que provee el DSP *Mediamath* el cual es el caso de estudio, se generaron y se elaboraron otros con transformaciones de los demás. Junto con ello, se validaron los atributos generados a través de herramientas estadísticas para confirmar que tienen un grado importante de explicatividad de la varianza de los datos. Es posible mejorar la etapa de generación de atributos adhiriendo los rubros de los clientes que planean hacer avisaje y alguna métrica que pueda evaluar la calidad del mensaje inscrito en las creatividades, lo segundo se deja propuesto dado que no se tiene la información y por lo tanto no se pudo agregar al modelo.

Luego de establecer los atributos relevantes para la predicción de las dos métricas de rendimiento de las campañas publicitarias, se elaboraron segmentos en las campañas debido a que hay dispersión y comportamiento heterogéneo entre ellos mismos, con el fin de comparar campañas con comportamiento similar y evitar calcular resultados para el promedio de estas. Los segmentos generados tienen interpretaciones intuitivas de su comportamiento y son realistas a modo de diferenciar entre ellas, es posible mejorar dicha segmentación utilizando otros métodos que el propuesto y agregando otras transformaciones de los atributos además características intrínsecas de los clientes y de los mensajes en las creatividades.

El último elemento del informe es la elaboración de los modelos predictivos y la valorización de ellos en términos monetarios. El resultado de la evaluación del modelo según las métricas del CTR y CPC respectivamente se muestra en las siguientes tablas:

Segmento	Margen Potencial [UM]
Publicidad Constante	-93,09
Promoción Premium	95,43
Experimentación	117,1

Tabla 24, Resumen de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo

Segmento	Margen Potencial [UM]
Publicidad Constante	-50,12
Promoción Premium	-127,04
Experimentación	-96,3

Tabla 2523, Resumen de los Costos de Rendimiento de las Campañas Publicitarias Reales vs Modelo

El margen potencial supone una comparación entre lo actualmente realizado por *Affiperf* en el mes de octubre en comparación a lo que predijo el modelo en el mismo periodo de tiempo. El modelo presenta resultados muy favorables en los segmentos de promoción premium y experimentación en comparación con los resultados reales, debido a que la predicción que se realiza actualmente proviene de la experiencia del encargado de negociación lo que claramente se puede mejorar. En el segmento de publicidad constante el modelo es inferior a las proyecciones utilizadas por *Affiperf* debido a que utilizan la información de las mismas campañas realizadas anteriormente para predecir la negociación, lo que supera al modelo. La recomendación es aplicar el modelo a las campañas que no tengan precedentes puesto a que supera evidentemente a los resultados actuales.

Por otro lado la predicción del CPC por parte del modelo es peor que la negociación realizada actualmente por *Affiperf* debido a que los atributos o variables explicativas utilizadas en el modelo no logran explicar la variabilidad del comportamiento de esta métrica, por lo que no se recomienda utilizarla en ninguno de los segmentos.

6.2 Recomendaciones a la Empresa

La metodología propuesta sugiere que se puede mejorar bastante la negociación con los clientes para campañas que no se posee información antigua. El hecho de que no se posea información genera que la opinión de expertos sea menos informada y por lo tanto pueda generar potenciales pérdidas o disminución de los márgenes. Se propone utilizar el modelo propuesto en las campañas publicitarias las campañas de corta duración y las que no tienen información histórica (segmentos 2 y 3) para la negociación y predicción del CTR.

Para la métrica del CPC no es concluyente que el modelo entregue un mejor resultado al que actualmente se está realizando, por lo que no se recomienda tomar decisiones basados en ello, pero puede ser ilustrativo e indicar un punto de referencia además de la opinión de expertos.

6.3 Trabajo Futuro

La metodología propuesta utiliza los datos directamente extraídos del DSP además de la generación de atributos a partir de los mismos. Es posible mejorar la capacidad descriptiva del modelo si se agregan métricas de calidad de las creatividades, debido a que podría tener un importante impacto en la conducta del cliente expuesto a la publicidad, en particular esta métrica podría explicar fuertemente la variabilidad del CPC permitiendo la aplicación de la metodología a este KPI de igual manera.

Por otro lado considerar los tipos de productos o servicios que se están publicitando podrían mejorar la capacidad predictiva dependiendo del grado de compromiso que se tiene con el producto. Es posible que productos de mayor compromiso como automóviles o habitaciones no sean tan efectivos en la publicidad como si los podrían ser los productos electrodomésticos por lo que se podría incorporar una métrica que considere este factor.

Finalmente se propone adaptar la metodología con la evolución del RTB, dado que tiene una constante evolución en la adquisición de información de los consumidores y esto podría generar nuevas oportunidades para expandir la metodología propuesta.

Glosario

- **Real Time Bidding (o compra programática):** Subasta en tiempo real por el espacio publicitario en internet.
- **Inventario:** Se refiere a los espacios en internet que están dispuestos para poner publicidad. Es dinámico y se activa cuando el usuario ingresa a un sitio con espacio para publicidad y desaparece en cuanto se cierra el mismo.
- **DSP (Demand Side Platform) :** Son las distintas plataformas que son configuradas para comprar inventario. Existen varias empresas que proveen las plataformas (Appnexus, Mediamath, StrikeAd, Dynadmic), en la que cada una provee diferentes algoritmos de compra . Son el objetivo del presente estudio.
- **Impresión:** Despliegue de la publicidad a un usuario en un sitio de internet habilitado para tal.
- **Creatividades:** Son los anuncios que serán impresionados al usuario. Estos pueden ser imágenes, videos, animaciones, etcétera.
- **Campañas Publicitarias:** Proyecto publicitario que desea comunicar la empresa.
- **Estrategias:** Cada campaña puede tener una o más estrategias para ser implementadas. A cada una de ellas se le asigna un presupuesto(en función del presupuesto de la campaña), y se le establecen variables de decisión para actual en la implementación.
- **Segmentación DSP:** La categorización de sitios, el tipo de dispositivo y el área geográfica que proveen los DSP para la implementación de campañas.
- **Exchange:** Conjuntos de sitios suministrados por un único proveedor para la venta, en otras palabras en una empresa que vende espacios publicitarios de un conjunto de sitios en la web.
- **Cookies:** Información adherida al navegador del usuario, la cual permite entre otras cosas, monitorear la navegación de la persona.
- **Navegador:** Programa que permite acceder a la web.
- **Plan de Medios:** Propuesta de diferentes medios para elaborar una campaña publicitaria.

Tipos de Compras: Existen diferentes tipos de compra en internet

- CPM: Costo por mil impresiones
- CPC: Costo por clic
- CPA: Costo por acción (acción en la web a definir por el anunciante)

Bibliografía

- [1] Panorama Medio Enero 2014, proyección 2015. ComScore
- [2] Subsecretaría de Telecomunicaciones, <http://www.subtel.gob.cl/penetracion-de-internet-en-chile-alcanza-los-642-accesos-por-cada-100-habitantes/>. Revisado 30 Octubre 2015.
- [3] Google. The arrival of real-time bidding, 2011.
- [4] Shuai Yuan, Jun Wang, Xiaoxue Zhao, Real-time Bidding for Online Advertising: Measurement and Analysis, 2013.
- [5] IN4521, Introducción a la Minería de Datos, Clase 4, primavera 2015.
- [6] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*", Berkeley, University of California Press, 1:281-297.
- [7] Davies, David L.; Bouldin, Donald W. (1979). "A Cluster Separation Measure". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PAMI-1 (2): 224–227.
- [8] R. Sibson (1973). SLINK: an optimally efficient algorithm for the single-link cluster method. *The Computer Journal* (British Computer Society) **16** (1): 30–34.]
- [9] Jonathon Shlens. A Tutorial on Principal Component Analysis. Revisado 27 Octubre 2015.
- [10] <http://www.havasmedia.com/press/press-releases>. Revisado 24 Octubre 2015.
- [11] <http://www.affiperf.com/about-us/company/>. Revisado 24 Octubre 2015.
- [12] Linear Regression Analysis, 2nd Edition. George A. F. Seber , Alan J. Lee.
- [13] http://www.financeformulas.net/Present_Value.html. Revisado 24 Octubre 2015.
- [14] AUC: a Better Measure than Accuracy in Comparing Learning Algorithms. Ling, Huang, Zhang. University of New Brunswick, Canada 2004.
- [15]
http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html,
revisado 5 Diciembre 2015.
- [16] <https://www.site.uottawa.ca/~stan/csi7162/presentations/William-presentation.pdf>,
revisado 5 Diciembre 2015
- [17] Comparing Evaluation Metrics for Sentence Boundary Detection. Liu and Shriberg, 2007

[18] https://en.wikipedia.org/wiki/Receiver_operating_characteristic. Revisado 5 Diciembre 2015

[19] AUC: a misleading measure of the performance of predictive distribution models. Lobo, Jimenez-Valverde y Real, Málaga 2007.

[20] [https://en.wikipedia.org/wiki/Lift_\(data_mining\)](https://en.wikipedia.org/wiki/Lift_(data_mining)). Revisado 5 Diciembre 2015

[21] https://en.wikipedia.org/wiki/Precision_and_recall. Revisado 5 Diciembre 2015

[22] IN-4521-1, Introducción a la Minería de Datos, clases 8-11. Primavera 2015

[23] [https://en.wikipedia.org/wiki/Cross-validation_\(statistics\)](https://en.wikipedia.org/wiki/Cross-validation_(statistics)). Revisado 5 Diciembre 2015.

Anexo y Apéndices

Anexo 1: Clusters Creatividades

Creatividad	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
creative_size1024x768	0.0	0.0	0.0	0.0	0.0222
creative_size120x600	1.0	0.1333	0.0112	0.0	1.0
creative_size160x600	1.0	0.5	0.022	0.944	0.8
creative_size168x28	0.0	0.066	0.0	0.0	0.0
creative_size200x200	0.109	0.0	0.0	0.0	0.0
creative_size216x36	0.0	0.1	0.0	0.0	0.022
creative_size250x250	0.109	0.0	0.0	0.0185	0.0
creative_size300x250	1.0	1.0	0.337	1.0	1.0
creative_size300x50	0.0	0.4666	0.0	0.0	0.066
creative_size300x600	1.0	0.0	0.0112	0.4814	0.0
creative_size320x48	0.0	0.0333	0.0	0.0	0.0
creative_size320x480	0.0	0.0333	0.0	0.0	0.0444
creative_size320x50	0.178	1.0	0.0	0.055	0.0
creative_size1024x768	0.0	0.0	0.0	0.0	0.02223

Anexo 2: Resultados Regresión sin Segmentación CTR

	Estimado	Desviación	Pr(> t)
(Intercept)	4,10E+02	1,47E+02	2.789
Cluster_Creatividades 1	-3,86E+02	2,43E+02	-1.585
Cluster_Creatividades 2	-5,20E+02	1,42E+02	-3.674
Cluster_Creatividades 3	-4,52E+02	1,38E+02	-3.271
total_spend_cpm	4,14E+02	4,49E+01	9.219
contextual_cost	5,92E-02	8,46E-01	0.070
Velocidad_Impressions	1,84E-02	1,24E-02	1.482
Velocidad_Spent	-3,58E+01	2,38E+01	-1.504
total_spend	-2,66E-01	1,73E-01	-1.533
Duracion	-2,66E-02	1,83E-02	-1.457
impressions	4,71E-05	7,28E-05	0.647

Anexo 3: Resultados Regresión sin Segmentación CPC

	Estimado	Desviación	Pr(> t)
(Intercept)	0,57	0,158	0,000353
Cluster_Creatividades 1	-0,063	0,262	0,81
Cluster_Creatividades 2	-0,2042	0,15	0,18
Cluster_Creatividades 3	0,1124	0,149	0,45
total_spend_cpm	0,002	0,048	0,96
contextual_cost	0,00047	9,132e-04	0,604
Velocidad_Impressions	1,298e-06	1,338e-05	0,92

Velocidad_Spent	-5,046e-03	0,026	0,855
total_spend	1,980e-04	1,870e-04	0,29
Duracion	-2,294e-05	1,970e-05	0,245
impressions	-1,519e-07	7,861e-08	0.054