



Especialización en Administración Bancaria

TRABAJO INTEGRADOR FINAL

“EFICACIA COMERCIAL: Venta direccionada de Plazos Fijos”

DOCENTE: DR. NELSON ROCA

ALUMNO: ING. GERARDO DARIO BROVEGLIO

AÑO 2009

INDICE

1 RESUMEN EJECUTIVO.....	3
2 PLANTEAMIENTO Y FORMULACIÓN DEL PROBLEMA DE INVESTIGACIÓN... 5	5
3 MARCO DE REFERENCIA CONCEPTUAL. CAMPAÑAS COMERCIALES..... 9	9
3.1 Introducción	9
3.2 Planificación y ejecución	9
3.3 Presupuesto	11
3.4 Factores que influyen en su éxito.....	12
3.5 Medición de la efectividad de una campaña comercial	13
4 ESTRATEGIAS DE MARKETING	15
4.1 Marketing Masivo.....	15
4.2 Cambiando de estrategia.....	16
4.3 Marketing Directo	16
4.4 Importancia del grupo objetivo	17
5 DETERMINACIÓN DEL UNIVERSO DE VARIABLES	19
5.1 Estrategias actuales de selección.....	19
5.1.1 Primera etapa: Fuentes primarias de recolección de datos. Filtrado	19
5.1.2 Segunda Etapa: Selección aleatoria	21
5.2 Conclusión de las estrategias actuales.....	22
5.3 Tasa de éxito actuales	23
5.4 Objetivo y Alcance del Proyecto.....	24
5.5 Mejora esperada	25
6 DISEÑO DE TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS... 27	27
6.1 Introducción	27
6.1.1 Data Mining	27
6.1.2 Orígenes del Data Mining.....	27
6.2 Modelos de Respuesta.....	28
6.3 Modelos aplicados a Marketing Directo	28

6.4	Proceso de Data Mining	30
7	DESARROLLO DE LA SOLUCIÓN	31
7.1	Requisitos	31
7.1.1	Software	31
7.1.2	Hardware	32
7.1.3	Datos	32
7.2	Etapas de Data Mining	32
7.2.1	Selección gruesa de variables	32
7.2.2	Pre-Procesado de Datos	35
7.2.2.1	Limpieza de Datos	36
7.2.2.2	Integración de Datos	36
7.2.2.3	Normalización de Datos	36
7.2.3	Selección Fina de Variables	37
7.2.3.1	Prelavado	37
7.2.3.2	Selección de Variables	38
7.2.4	Transformación	39
7.2.5	Creación de Modelos	40
7.2.6	Combinación de modelos parciales para obtener el modelo final	42
8	IMPLEMENTACIÓN Y ANÁLISIS DE RESULTADOS	43
8.1	Prueba del Modelo en caso Real	43
8.2	Impacto económico logrado	43
8.2.1	Ingresos Generados	43
8.2.2	Costos incurridos	45
8.2.2.1	Software	45
8.2.2.2	Hardware	45
8.2.2.3	Mano de Obra	45
8.2.3	Impacto económico	46
9	CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	47
10	BIBLIOGRAFÍA	49

1 Resumen Ejecutivo

El entorno que nos rodea moldea nuestro parecer, modifica nuestro punto de vista, condiciona nuestro comportamiento. Interactuando con otros individuos compartimos opiniones e imitamos conductas. Con la aparición de los sistemas informáticos, la conducta del hombre comenzó a ser registrada y almacenada en forma de datos en las cada vez más grandes memorias de las computadoras. Tanto información personal como aquella que, al igual que una huella, generamos en nuestro paso por algún comercio, esta disponible para ser utilizada. Miles de patrones de comportamiento subyacen en esa masa de datos. Con las técnicas adecuadas es posible encontrarlos. Podemos anticipar la respuesta de un individuo con tan solo conocer sus datos personales y la forma en que se ha comportado en el pasado.

Hace algunos años, un grupo de profesionales entre los que se encontraban estadistas, ingenieros, sociólogos, entre otros, se reunió para discutir acerca de estos temas. De aquella reunión resultaría lo que fue la primera versión de lo que actualmente se conoce bajo el nombre de "Data Mining". El objetivo de estas técnicas consiste en "minar" los datos en busca de patrones de conducta de los individuos y que, al encontrarlos, permita mejorar la toma de decisiones en las empresas.

Mediante el presente trabajo se intentará obtener un modelo predictivo de respuesta para la cartera de clientes de un Banco.

Para hacerlo, se aplicará Data Mining sobre la base de datos de dichos clientes y como resultado se obtendrá un modelo que asocie a cada individuo la probabilidad de adquirir el producto que se le ofrece. Contando a priori con esta información, las ofertas se realizarán a aquellos clientes que tengan mayor propensión a aceptar, según nuestro modelo, y se descartarán aquellos de los que no se esperen resultados positivos. Como consecuencia, se aumentará la tasa de éxito de las campañas comerciales, aumentando la colocación de, en este caso, nuestro foco serán los Plazos Fijos.

2 Planteamiento y formulación del problema de investigación

Para impulsar los niveles de ventas, las empresas implementan estrategias de marketing cuyo objetivo es promocionar sus productos o servicios entre el público en general. Mediante esta acción, se busca que quien recibe el mensaje se sienta atraído por el producto y, en consecuencia, realice la compra.

Dos aspectos principales son los que caracterizan a estas acciones de promoción que, en los Bancos, se denominan Campañas Comerciales y cuyo objetivo es vender la mayor cantidad posible de productos o servicios.

El primero de ellos tiene que ver con el **alcance**, es decir, el número de personas a quien se quiere enviar el mensaje. A ese grupo se lo llama grupo objetivo (target-group).

El segundo aspecto está relacionado con la **calidad del grupo objetivo**. Es decir, la respuesta obtenida. Por ejemplo, podemos comunicar a todas las mujeres del país que nuestra empresa ha sacado al mercado un nuevo modelo de taladro rotopercutor y sin embargo el nivel de respuesta será lo suficientemente bajo como para considerar la campaña un fracaso. Si, en cambio, el grupo objetivo hubiesen sido hombres de entre 25 y 45 años, se puede afirmar casi con seguridad que se habrían logrado mejores resultados.

Ambos aspectos son determinantes en el éxito de las campañas y de su combinación surgen los niveles de venta finales. Según el alcance definido y cuanto se busque conocer el grupo objetivo, se estarán utilizando una u otra estrategia de marketing.

Si bien el alcance elegido dependerá del presupuesto disponible para las campañas y a su vez del mercado real que se estime para el producto o servicio, la calidad del grupo objetivo siempre es deseable y cuanto mejor sea, mejores resultados se obtendrán. No obstante esto, cabe aclarar que lograr dicho conocimiento tiene un costo asociado, el cuál, utilizando Data Mining, es lo suficientemente bajo como para ser una alternativa irrefutable.

Para lograr un alto nivel de respuesta es necesario segmentar, o sea, conocer el universo de clientes, estudiarlo, identificar cada individuo según sus patrones de comportamiento y luego determinar a qué grupo se quiere dirigir el esfuerzo de venta.

Las publicidades tradicionales tal como las que pueden verse en la vía pública o en programas de televisión son publicidades masivas y se caracterizan por un gran alcance y una baja segmentación. Es decir, comunican a un gran número de personas aunque no identifican de manera individual a los receptores del mensaje.

Es fácil entender que, para lograr que una estrategia tenga posibilidades de ser exitosa, debe asegurar al menos uno de los dos aspectos que las caracterizan: un gran alcance o, ante un alcance reducido, una alta respuesta que asegure el nivel de ventas esperado, pero ninguno de estos dos casos se da en las estrategias comerciales del Banco en estudio.

En particular se tratará el caso de las campañas de Plazo Fijo donde el volumen de clientes alcanzado no supera el 1% del total de la cartera que, en nuestro análisis, el universo será de 200.000 clientes.

Del total de clientes se realizan una serie de filtros que buscan asegurar en el grupo objetivo algunas características mínimas consideradas de importancia para lograr una buena respuesta en los clientes. Por ejemplo: "El cliente tiene que pertenecer al NSE¹ A, B o C y ser mayor de 21 años".

Al filtrar según estos criterios, se obtiene una base de 100.000 clientes que aún supera ampliamente el 1% definido como alcance. Por ende es necesario realizar una nueva selección sobre este grupo. ¿Cual es la estrategia utilizada para hacerlo? En forma aleatoria.

La situación actual es la siguiente: "Las tasas de éxito resultantes están por debajo de las esperadas"

¹ NSE: Nivel Socio Económico se basa en los siguientes indicadores:
Nivel Educativo del Principal Sostén del Hogar (indicador de mayor importancia)
Nivel Ocupacional del Principal Sostén del Hogar
Patrimonio del Hogar (Posesión de bienes y de automóvil)

http://cmapspublic3.ihmc.us/rid=1194899212562_2140868026_3967/DEFINICION%20DEL%20INDICE%20DE%20NIVEL%20SOCIO.pdf

Ante esta situación cabe preguntarse lo siguiente.

¿Es esta la mejor tasa de respuesta que puedo obtener?

¿Todos los clientes del Banco tienen la misma propensión a adquirir el producto?

¿Es posible saber algo acerca de la propensión de unos y otros a adquirir el producto ofrecido?

La respuesta para la primera pregunta es: Probablemente sí, mientras se siga utilizando la selección aleatoria de clientes.

La respuesta para la segunda pregunta es No. Cada persona tiene características personales y patrones de comportamiento observables que lo hacen más propenso a adquirir cierto producto o servicio. Por ejemplo: un hombre es más propenso a adquirir un taladro que una mujer.

Respecto de la tercera pregunta, la respuesta es Sí. Es posible hacerlo y mediante este trabajo se busca mostrar en que forma es posible aumentar el éxito de las campañas comerciales simplemente mediante la selección inteligente del grupo objetivo. Es decir, eligiendo mejor a quien vender.

Para hacerlo se buscará asociar a cada cliente con la propensión que tiene de adquirir el producto ofrecido en determinada campaña. La selección ya no se hará mediante filtros o en forma aleatoria sino que se elegirá a quienes presenten una mayor probabilidad de adquirir el producto.

Como resultado se espera aumentar la tasa de éxito de las campañas y por ende aumentar el beneficio económico percibido por el Banco.

3 Marco de referencia Conceptual. Campañas Comerciales

3.1 Introducción

La actividad principal del Banco consiste en tomar dinero pagando a cambio cierta tasa de interés pasiva y luego prestar parte de ese dinero cobrando a cambio una tasa activa mayor.

Tanto para prestar como para captar dinero, el Banco utiliza distintos productos ya sean pasivos o activos. Mediante dichos productos, un cliente puede efectuar un depósito (Caja de Ahorro, Cuenta Corriente, Plazo Fijo) o pedir prestado dinero (Préstamos).

Si bien los clientes solicitan estos productos por iniciativa propia, el Banco encara acciones de promoción para impulsar sus ventas y de esa forma incrementar las operaciones activas y pasivas para luego, mediante el margen de intermediación, obtener mayores ganancias.

Estas acciones tienen el nombre de Campañas Comerciales y tanto en la definición como en la posterior implementación, se involucra a distintas áreas del Banco entre las cuáles están las áreas de Marketing y de Inteligencia Comercial. Cuando las campañas están dirigidas a clientes existentes a los cuales se busca venderles productos adicionales a los que actualmente poseen, se las denomina campañas de cross-selling, las cuales son destinadas a incrementar el número de clientes.

Los canales elegidos usualmente para contactarse con los clientes son vía e-mail, teléfono, a través de las sucursales o, en ocasiones, directamente a través de los ATMs (cajeros electrónicos) dependiendo del producto del que se trate.

3.2 Planificación y ejecución

Cuando se quiere salir al mercado con una campaña comercial, es necesario definir: el **producto** a vender, el **canal** mediante el cual se va a vender, el

período durante el cual se desarrolla la campaña y por último **a quién** se quiere vender (ver Ilustración 1).



Ilustración 1: Variables de control de las Campañas Comerciales

PRODUCTO OFRECIDO: Es el producto que se desea vender en la campaña; dependerá de las necesidades del Banco en cada momento.

Puede responder a una necesidad financiera, es decir, aumentar la captación o la colocación de fondos de cierto tipo o a la fidelización de clientes mediante el ofrecimiento de productos que el mercado valora en ese momento.

Los atributos del producto (tasa de interés, costos de mantenimiento y disponibilidad de fondos) se eligen en base a la rentabilidad esperada, en base a las necesidades del mercado y a su vez a través de benchmarking (análisis de la competencia)

CANALES DE VENTAS: Se trata del punto de contacto elegido para comunicarse con los clientes: sucursales, telefónico, cajeros automáticos, banca telefónica, o por carta, entre otros. En general se refuerza la acción de

los canales de venta con folletos informativos que deben ser diseñados, impresos y distribuidos.

Para determinar que canal se utilizará en una campaña comercial se tienen en cuenta dos cuestiones. 1º: Eficiencia de cada una de ellos en campañas pasadas. En promedio, algunos canales obtienen mejores resultados que otros. 2º: Disponibilidad que presentan al momento de lanzar la campaña. Las sucursales, por ejemplo, están abocadas a otras tareas como pueden ser campañas lanzadas previamente, con lo cual habría que optar en ese caso por utilizar la Banca Telefónica o los cajeros automáticos.

PERIODO: El período es simplemente el espacio de tiempo durante el cual estará vigente la campaña. El contexto económico, político y social es un factor importante a considerar dado que la venta de un producto esta íntimamente relacionada con las necesidades de las personas en un momento dado.

GRUPO OBJETIVO: Son aquellos clientes a quienes se contactará para intentarles vender el producto. De su decisión depende el éxito o el fracaso de la campaña. Al igual que la resolución de un juicio depende del jurado elegido, las ventas resultantes dependen del grupo objetivo seleccionado.

3.3 Presupuesto

Las campañas comerciales tienen costo para las entidades financieras, las mismas dependen de las características de la campaña pero en general los costos se dividen en costos de diseño, impresión y distribución (si se envían a domicilio) de los folletos que se entregan a los clientes como refuerzo del contacto directo. Si la campaña es telefónica y se realiza mediante un proveedor externo también es necesario pagar por dicho servicio.

El costo de horas hombre se considera un costo hundido ya que los sueldos fijos de los empleados se pagarán se realice o no la campaña. No obstante es necesario considerar la disponibilidad de horas hombre de cada canal (sucursal, call center interno, fuerza de ventas, entre otros) las cuales también limitan el volumen de clientes a contactar.

El Banco asigna una parte del presupuesto a las Campañas Comerciales, el cuál limita tanto los costos como las horas hombre disponibles.

Esto implica que no se puede contactar al 100% de los clientes.

Una vez definido el presupuesto y en función de los costos, queda determinado la cantidad de contactos que se pueden realizar en cada campaña.

Una campaña comercial contacta, en promedio, 3000 clientes.

3.4 Factores que influyen en su éxito

Para mejorar el éxito de una campaña comercial, se puede intervenir en cualquiera de los 4 puntos mencionados anteriormente: Producto ofrecido (sus atributos), Canal de venta (pieza de marketing), Período (contexto) en el cual se lanza la campaña y por último grupo objetivo (grupo de clientes a contactar). Dicha relación se resume en la siguiente fórmula:

$$V(x) = f(\text{Producto, Canal, Periodo, Grupo Objetivo}) \quad \mathbf{V(x): Ventas}$$

Mejorando los atributos del producto (desde el punto de vista del cliente) puede incrementarse la demanda. Por ejemplo, ofreciendo una mejor tasa de interés.

Eligiendo canales de contacto directos (sucursales) puede lograrse una mayor respuesta que si solo se envía una carta al domicilio del cliente.

Lanzando la campaña en el momento adecuado, pueden aprovecharse situaciones particulares como, por ejemplo, la necesidad de préstamos hipotecarios dada la escasez de los mismos en el mercado.

Por último, el grupo objetivo se vuelve importante al existir una restricción en el alcance de la campaña y por ende, ser necesaria la selección de cierto grupo de clientes a contactar entre el total de la cartera. No todos los clientes responden de igual forma a las ofertas. A mayor respuesta, mayor volumen de ventas.

3.5 Medición de la efectividad de una campaña comercial

Si bien las campañas son un medio mediante el cual se busca ayudar al Banco a lograr sus objetivos, es decir, lograr cierto nivel de ingresos través de la toma y préstamo de dinero, las mismas tienen objetivos propios.

El indicador más importante de la efectividad de una campaña comercial es la tasa de éxito (hit rate) que resulta de la división matemática entre el número de ventas efectivas y el número de clientes contactados, expresada en porcentaje.

El valor de este indicador al término de la campaña indica en que medida se alcanzó el objetivo propuesto. El máximo valor posible corresponde al 100% que implica una situación ideal donde todos y cada uno de los clientes contactados adquieren el producto. Si bien es improbable alcanzar esa efectividad, en cada campaña se intenta obtener un valor lo más cercano posible a él.

El costo de contacto de cada cliente es significativamente inferior al ingreso que resulta de la venta efectiva. Si se gastara más dinero en convencer a un cliente que el esperado como ingreso al lograr la venta, no existe posibilidad alguna de obtener beneficios, aún logrando el ideal de 100% en la tasa de éxito.

A continuación se muestra la ecuación para el cálculo de la tasa de éxito y el beneficio obtenido a raíz de la campaña.

$$E(\%) = V / C$$

$$B = V * Iu - C * Cu$$

E(%): Tasa de Éxito
V: Ventas (cantidad de clientes)
C: Clientes Contactados
Iu: Ingreso unitario
Cu: Costo unitario del contacto
B: Beneficio de la campaña

Igualando $B = 0$ es fácil ver que la tasa de éxito mínima debe ser superior a la relación entre el costo unitario del contacto y el ingreso unitario, tal como muestra la siguiente ecuación:

Para $B > 0$	$E > Cu / Iu$
--------------	---------------

Aumentar las tasas de éxito es equivalente a aumentar las ventas y por ende a incrementar el ingreso obtenido por el Banco, siendo dicha relación directa.

No existen costos fijos a considerar dado que todo gasto depende del volumen de clientes que se contacta.

Para $R > 0$
 $E > C_{fijo}$

4 Estrategias de Marketing

4.1 Marketing Masivo

El marketing masivo es una estrategia de venta mediante la cuál una empresa decide ignorar las diferencias entre los distintos segmentos de clientes (personas agrupadas según sus características) yendo detrás de todo el mercado con un único mensaje.

La hipótesis en la cuál se basa esta estrategia es que alcanzando la mayor audiencia posible se maximiza la exposición al producto y por ende como resultado se obtendrá un mayor número de ventas.

Un claro ejemplo son las publicidades a través de radio o televisión.

Sin embargo desde algún tiempo esta forma de llegar a los consumidores ya no logra ser tan efectiva como lo fue durante la mayor parte del siglo XX.

Según una encuesta realizada por Yankelovich Partners Inc², el 65% de los consumidores se sienten constantemente bombardeados con demasiadas publicidades. A su vez, el 61% cree que el volumen de avisos esta fuera de control.

El presidente de esta firma, J. Walker Smith, agrega: "El problema es que actualmente a los encargados de marketing no les preocupa la imprecisión, sin darse cuenta del impacto que eso tiene en aumentar la resistencia y reducir la productividad del marketing."³

En línea con lo anterior, en un artículo titulado "Cambiando de Marketing Masivo a Micro Marketing", Ian Durrell Líder de Desarrollo de Negocios de Only Finance Ltd., expone: "Para satisfacer a los consumidores, las empresas deben pasar de transmitir mensajes masivos a un gran número de personas a alcanzar de forma individual a cada uno de ellos"⁴

² <http://www.yankelovich.com/>

³ <http://www.clickz.com/3344701> - Consumers Becoming Marketing-Resistant

⁴ Shifting From Mass Marketing to Micro Marketing
<http://www.marketingprofs.com/5/durrell1.asp>

4.2 Cambiando de estrategia

El comportamiento del mercado ha cambiado. Actualmente la diversidad en las ofertas es tan grande que frente a tantos avisos, las publicidades pierden impulso para tratar de instalarse en la mente de los consumidores. En consecuencia, solo una pequeña porción de los consumidores adquirirá el producto o servicio.

El marketing de segmentos no es una novedad. Se trata de identificar dentro del total de clientes potenciales, aquellos que cumplan con ciertas características (segmento) definidas según el producto para luego dirigir la publicidad a ese segmento en particular.

Si bien el nivel de segmentación es mayor que en el marketing masivo, el marketing de segmentos también ha dejado de ser una estrategia novedosa y de gran utilidad.

En un artículo denominado "Uno por Uno", Horacio Marchand⁵ escribe: "Primero fue atender clientes masivamente y de forma anónima. Luego apareció la segmentación y se atendía a grupos específicos de clientes. Hoy se busca atender un cliente a la vez y generar conocimiento preciso de sus hábitos y patrones de consumo. El Segmento de Uno se está imponiendo."

El autor resalta en este párrafo la importancia de contactar a los clientes Uno a Uno, es decir, en forma individual.

4.3 Marketing Directo

Se llama Marketing Directo a la estrategia de venta mediante la cual se contacta a los clientes en forma individual ya sea por teléfono, e-mail o en persona.

Es un enfoque opuesto al Marketing Masivo, dado que en lugar de enviar un único mensaje a toda la audiencia de potenciales clientes, se busca lograr

⁵ Licenciado en Adm. de Empresas y MBA en Austin, Texas

contactos individuales con aquellos clientes que sean más propensos a aceptar la oferta.

No obstante, conocer quienes son los clientes que con mayor probabilidad aceptarán la oferta implica conocer sus características y necesidades.

Lo que se espera es lograr un mayor nivel de ventas sobre los clientes contactados.

4.4 Importancia del grupo objetivo

Las campañas comerciales utilizan marketing directo como estrategia de ventas. Esto puede hacerse mediante el envío de e-mails, en el caso de campañas vía Internet, realizando llamadas telefónicas o dirigiéndose a ellos vía correo postal.

En los casos mencionados, es necesario tener en cuenta dos aspectos: el costo de contactar a todos los clientes y el efecto negativo que puede provocar contactar a clientes que no quieren ser contactados.

En el primer caso es necesario reducir los costos. Contactar a todos los clientes de la cartera mediante correo postal o llamadas telefónicas son tan elevados que superan la ganancia esperada ya que solo una pequeña porción aceptará la oferta.

En el segundo caso, es importante tener en cuenta que llamadas o folletos indeseados logran un efecto negativo que reduce el mercado potencial para futuras acciones de venta.

Además de los costos y el efecto negativo de contactos indeseados, las horas hombre disponibles, por ejemplo, en el caso de contactos telefónicos, es limitada.

En función de lo dicho anteriormente, se vuelve necesario seleccionar subgrupos de clientes para cada campaña.

Enfocar la campaña en aquellos clientes que son más propensos a adquirir el producto es un aspecto determinante para el éxito.

La adecuada selección del grupo objetivo se vuelve, entonces, imprescindible.

5 Determinación del universo de variables

5.1 Estrategias actuales de selección

Como se explicó anteriormente, que haya que elegir un sub-grupo de clientes a quien ofrecer los productos hace necesario definir los criterios en los cuales se basará dicha selección buscando al mismo tiempo que el grupo resultante tenga una mayor propensión a adquirir el producto.

Dado lo observado en las estrategias actuales de selección del target de sus campañas de ventas, el Banco no está logrando cumplir con lo mencionado anteriormente.

Para poder entender en forma más detallada el problema, se hará una descripción del proceso de selección del grupo objetivo que tiene lugar en forma previa al lanzamiento de una campaña.

La selección tiene 2 etapas.

5.1.1 Primera etapa: Fuentes primarias de recolección de datos. Filtrado

Al comienzo del proceso de selección, se cuenta con el total de clientes de la cartera, dicho total podemos estimarlo por ejemplo en los 200.000 individuos. El total de clientes se representa en la ilustración 2 donde A implica un cliente que, para cierto producto, canal y período, aceptaría la oferta comercial. Por otro lado, los N resultarían en contactos negativos.

Total de Clientes

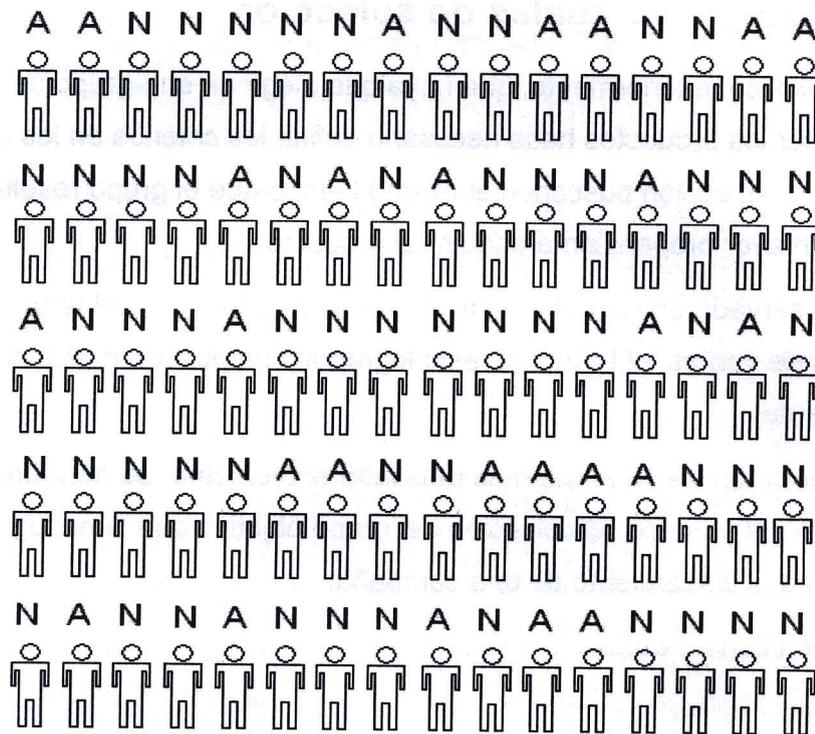


Ilustración 2: Total de Clientes (A: aceptará, N: No aceptará)

Para reducir el número de clientes a considerar para un potencial contacto, se realiza un primer filtro donde, se seleccionan aquellos clientes que cumplan las condiciones del tipo: “personas mayores a 30 años”, “NSE A o B excluyendo C”, “antigüedad mayor a dos años en el Banco”.

Habiendo concluido este primer paso, la base se ha reducido de 200.000 clientes a un total que, en general, se ubica en 100.000 clientes (ver ilustración 3), siendo este un número aún elevado considerando que se contactarán aproximadamente 3.000.

Cientes que cumplen con las condiciones mínimas definidas

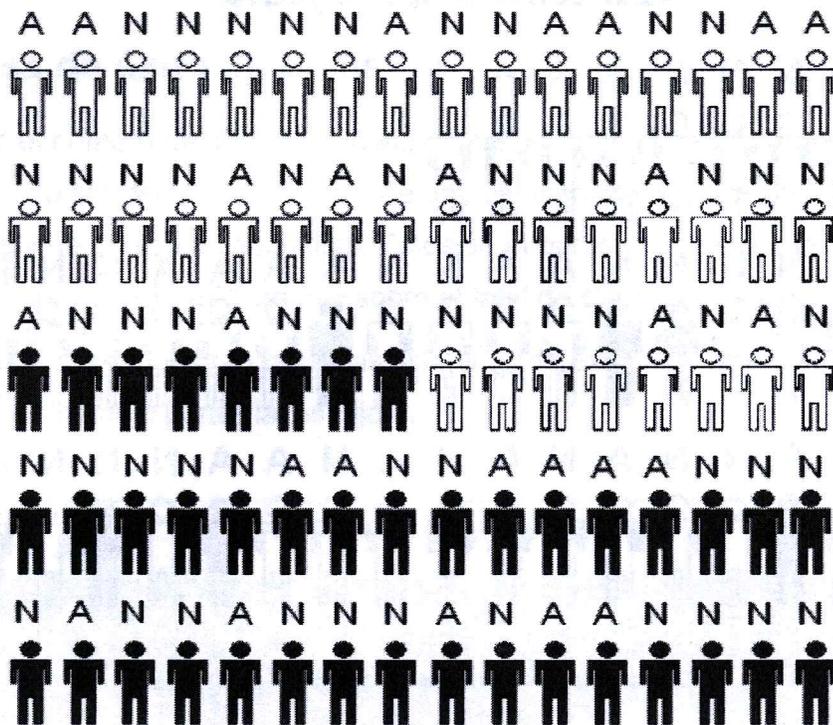


Ilustración 3: Clientes seleccionados en primer filtro

El problema que se detecta en esta primera etapa es que los criterios de selección mencionados se definen en base a reglas preestablecidas y de las cuales no se ha comprobado que realmente estén relacionadas con la propensión de los clientes a aceptar el producto. Esto pone en duda si realmente, los individuos elegidos (color verde en el gráfico) realmente tienen mayor propensión a adquirir el producto que los dejados fuera (color blanco en el gráfico) Las mismas reglas se aplican en todos los casos.

5.1.2 Segunda Etapa: Selección aleatoria

La segunda parte consiste simplemente en tomar el grupo resultante del filtro inicial y realizar sobre dicho grupo una selección aleatoria.

Es decir, partiendo de 100.000 clientes, se eligen 3000 individuos *al azar*.

Individuos elegidos al azar como Grupo objetivo

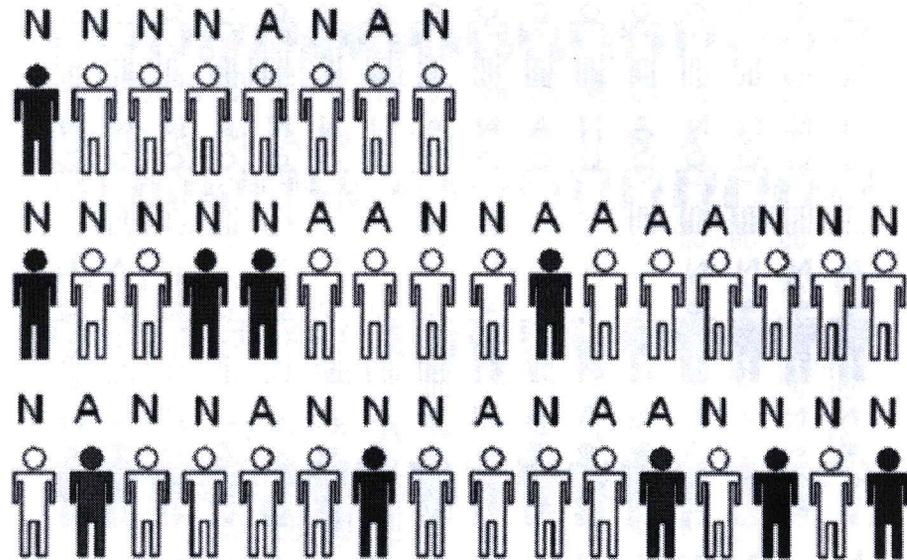


Ilustración 4: Clientes seleccionados al azar

Estos individuos forman, finalmente, el grupo objetivo a contactar.

Se muestran de verde en la Ilustración 4, aquellos clientes que fueron elegidos y finalmente aceptaron la oferta, mientras que aquellos de color rojo fueron elegidos pero no aceptaron la oferta.

El problema aquí detectado es, justamente, la selección aleatoria la cual no distingue entre las distintas características de los clientes y los considera igualmente atractivos. La variabilidad encontrada en una muestra de 100.000 clientes es tan alta que tomarlos como semejantes demuestra una ineficiencia en la selección.

5.2 Conclusión de las estrategias actuales

En ambas etapas se encontraron puntos débiles en la estrategia de selección, ya que en la primera etapa, se trató de la falta de fundamentos en los criterios utilizados, mientras que, en la segunda etapa, la selección aleatoria no agrega valor a la misma.

Si fuese posible refinar la selección de los clientes a contactar en función de criterios probados bajo rigor estadístico y aplicable al total de la cartera, sería

posible distinguir aquellos clientes que tienen mayor propensión a adquirir el producto y así aumentar el éxito en las ventas de las campañas comerciales.

5.3 Tasa de éxito actuales

Dado que el número de clientes contactados es distinto en cada campaña, cuando se quiere determinar el éxito de una campaña se lo hace en términos de hit-rate y no de ventas absolutas. Es decir, el porcentaje de clientes que realizaron un depósito a Plazo Fijo sobre el total de clientes contactados.

A continuación se muestran las tasas históricas de éxito de campañas de venta de Banco en estudio para Plazos Fijos⁶:

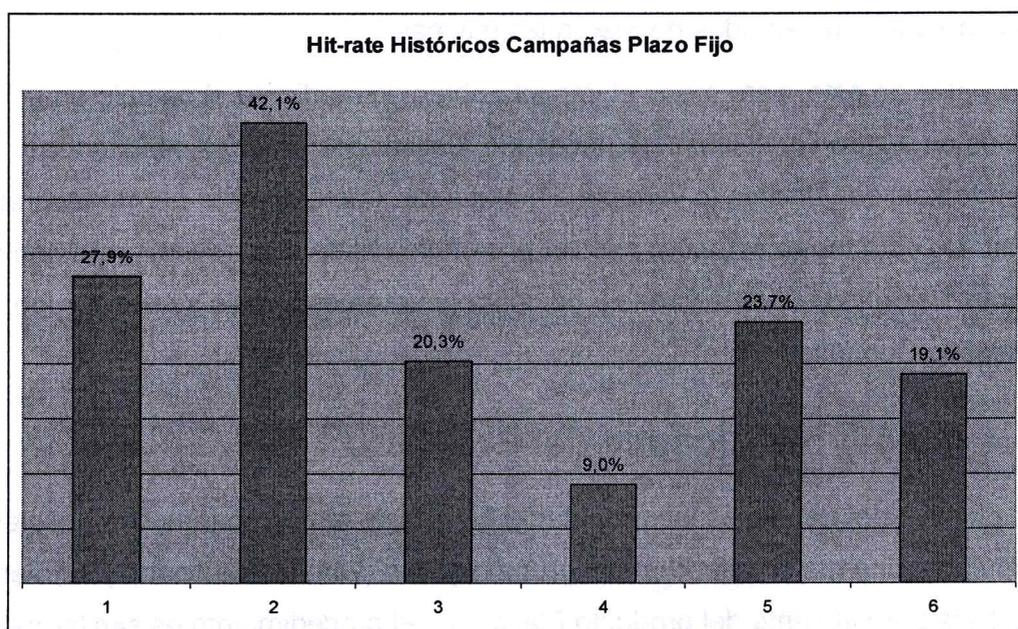


Gráfico 1: Tasas de éxito en campañas de Plazos Fijos

El hit-rate promedio simple es de 23.7%. Es decir que, en promedio, de cada 100 clientes contactados, casi 24 han aceptado la oferta.

⁶ Datos históricos Campañas Comerciales Plazos Fijos (ejemplo de una entidad financiera en plaza)

No obstante, a la hora de evaluar el impacto de la solución propuesta más adelante, este número no será representativo dada la heterogeneidad entre las distintas campañas.

Como se explicó en el capítulo 2, "Campañas Comerciales", existen cuatro factores que determinan el grado de éxito de una campaña comercial: Producto ofrecido, canal de ventas, período y grupo objetivo contactado

Las tasas de éxito observables en el gráfico anterior corresponden a campañas donde los productos ofrecidos en cada una no poseían las mismas características (distintas tasas de colocación), fueron ofrecidos en distintos momentos con lo cuál el contexto económico y social no fue el mismo, se utilizaron distintos canales influyendo en el éxito de cada contacto y por último las personas contactadas no fueron las mismas.

En vistas de esto, para medir el impacto obtenido mediante el cambio en una de estas cuatro variables, será necesario realizar dos pruebas simultáneas donde solo sea distinta la variable a monitorear. Las otras tres permanecerán fijas. De esta forma se podrá asociar el cambio observado en el hit-rate con el cambio ejercido sobre la variable de control que en este caso será el Grupo Objetivo Contactado.

5.4 Objetivo y Alcance del Proyecto

El objetivo del trabajo es incrementar el hit-rate de las campañas comerciales mediante una mejor selección del grupo objetivo (variable de control). Si bien se hará para la venta del producto Plazo Fijo, el procedimiento es similar para Préstamos, Tarjetas de Crédito, entre otros.

No se estudiará el impacto del canal de contacto ni se ahondará en la definición de los productos a ofrecer. Tampoco se estudiará el efecto del contexto en cada caso.

Lo que se quiere lograr es que, una vez determinados el producto, el canal y el período poder determinar el grupo objetivo que mayor propensión tiene a aceptar la oferta.

Para medir el efecto logrado mediante la solución propuesta, se medirán dos campañas simultáneas: una seleccionando el target group tal como se hace actualmente y la segunda según el procedimiento propuesto como solución.

5.5 Mejora esperada

Se espera poder aumentar el hit-rate en forma significativa logrando una mejora de al menos 15% entre ambas estrategias.

6 Diseño de Técnicas e Instrumentos de recolección de datos

6.1 Introducción

6.1.1 Data Mining

Data Mining es el proceso que permite obtener información justa y necesaria a efectos de los objetivos. Durante ese proceso se utiliza un conjunto de técnicas de análisis y modelización que son utilizadas según la pregunta a resolver.

La forma en que se logra esto es revelando relaciones implícitas, tendencias, patrones que estaban ocultos previamente para el ojo de los analistas.

Como dijo Einstein una vez: "Hacer la pregunta correcta es más que la mitad de la respuesta". Data Mining ayuda al analista de negocios a hacer las preguntas correctas para resolver problemas.

6.1.2 Orígenes del Data Mining

Las técnicas de Data Mining son el resultado de un largo proceso de investigación y desarrollo. Los componentes principales del Data Mining han evolucionado a lo largo de muchas décadas en áreas de investigación como la estadística, inteligencia artificial y aprendizaje computacional.

Las técnicas de Data Mining (redes neuronales, reglas inductivas, entre otros) han sido utilizadas durante muchos años para el desarrollo de sistemas de reconocimiento de patrones, reconocimiento óptico de caracteres, aplicaciones científicas y hasta en las Bolsas de Comercio.

El desarrollo de softwares de aplicación ha hecho hincapié en aumentar la facilidad de utilización buscando aumentar el mercado potencial, a la vez que se desarrollaban métodos de análisis más eficientes.

Softwares con interfases amigables, desarrollo tecnológico global y almacenamiento de datos, son los factores que hacen posible y práctica la utilización de Data Mining en una amplia variedad de rubros logrando obtener grandes beneficios por su aplicación.

6.2 Modelos de Respuesta

Una de las aplicaciones que ofrece el Data Mining, es la creación de “Modelos Predictivos de Respuesta”.

Al igual que utilizando regresión lineal pueden predecirse resultados futuros en función de una o más variables conocidas, los modelos de respuesta permiten anticipar el resultado futuro de acciones sobre clientes, en este caso, la acción es la oferta de un producto mediante contacto directo.

La forma que tienen estos modelos de predecir el resultado es encontrando, a partir de casos conocidos, cuales son las características comunes a aquellos clientes que han adquirido el producto. Aquellos atributos que resulten fuertemente relacionados con el resultado que se quiere predecir, serán considerados variables predictivas para el modelo.

6.3 Modelos aplicados a Marketing Directo

Como se explicó en el punto anterior, los modelos de respuesta permiten conocer, en función de casos conocidos, la propensión que tiene cada cliente a aceptar una oferta.

Aplicando el modelo a toda la muestra, puede establecerse quienes son aquellos clientes que con mayor probabilidad “comprarán” el producto. En consecuencia se logrará la selección de un grupo objetivo mediante técnicas de Data Mining con fundamentos estadísticos, procurando lograr el mayor nivel de ventas posible. El modelo es la solución buscada para el problema mencionado en el capítulo 4. Es una herramienta complementaria al Marketing Directo que mejora la eficiencia de esta estrategia.

A continuación se describe la estrategia propuesta de selección del grupo objetivo utilizando modelos de respuesta logrados con Data Mining.

Lo que se logra mediante esta estrategia es asignar a cada cliente, mediante el modelo de respuesta, la probabilidad esperada de que, luego del contacto, realice un depósito de Plazo Fijo. El esquema que ilustra el procedimiento es el siguiente:

Grupo objetivo elegido en base al Modelo de Respuesta

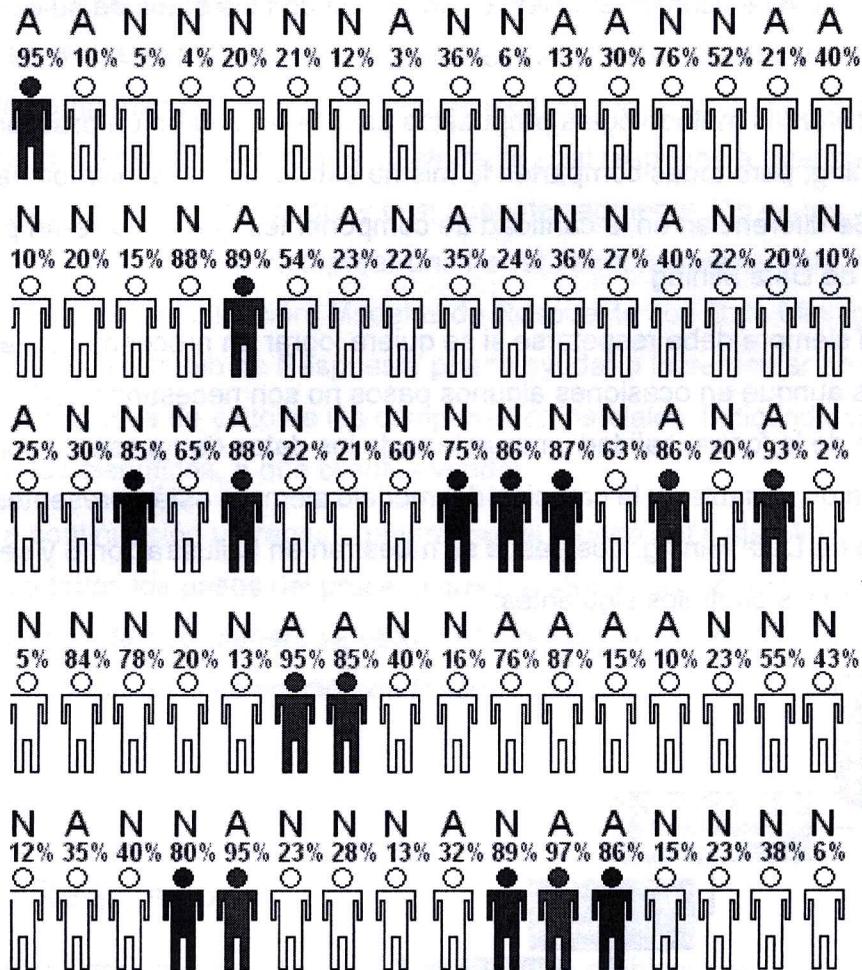


Ilustración 5: Clientes seleccionados con Modelo de Respuesta

Como se ve en la ilustración anterior, los clientes elegidos para el contacto son aquellos con las probabilidades de éxito más altas acorde al modelo.

Dada la incertidumbre inherente a las decisiones humanas, existirán clientes que rechazarán la oferta (marcados de color rojo).

6.4 Proceso de Data Mining

El proceso de Data Mining es un proceso estándar que se aplica en forma secuencial hasta alcanzar el objetivo. Comienza con los datos en su estado original y culmina con la creación del modelo.

Existen diversas metodologías propuestas para llevar a cabo un proceso de Data Mining, pero todas comparten la misma estructura y los mismos pasos a seguir. Se diferencian en la cantidad de componentes que consideran parte del proceso de Data Mining

El orden siempre debe respetarse si se quiere lograr un proceso ordenado y sin falencias aunque en ocasiones algunos pasos no son necesarios. Esto depende de la forma, calidad y naturaleza de los datos de que se dispone. La selección de variables y la creación del modelo siempre están presentes en un proyecto de Data Mining. Los pasos se muestran en la ilustración 6 y se detallan en los capítulos siguientes:

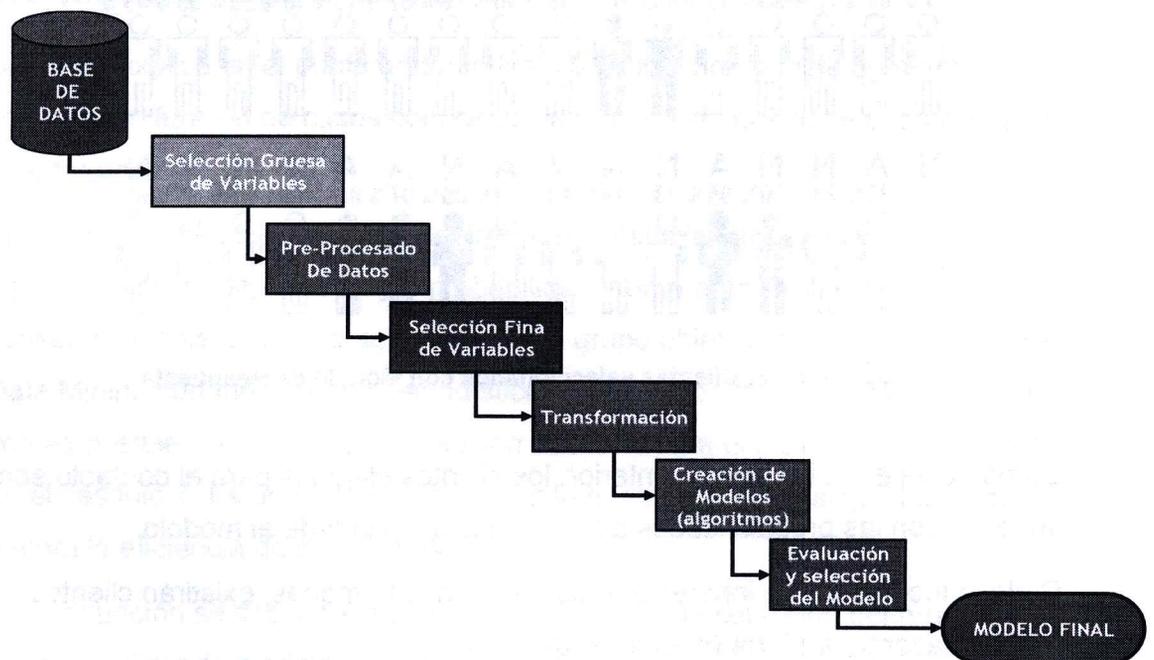


Ilustración 6: Proceso del Data Mining

7 Desarrollo de la solución

Hasta aquí ya se ha hecho una introducción al negocio bancario y se ha explicado la importancia que tiene, dentro de este, la colocación de Plazos Fijos. A su vez se explicó como el Banco favorece las ventas de dicho producto mediante campañas comerciales, luego se explicó que el grupo objetivo se determina prácticamente en forma aleatoria, lo cual responde a una estrategia de marketing masivo por la pobre y casi ausente segmentación de los clientes. Esto trajo la posibilidad de incorporar una herramienta novedosa para ser más inteligentes en dicha selección: Modelos de Respuesta con Data Mining. Se explicó como un Modelo de Respuesta puede ayudar a incrementar en forma sustancial las tasas de éxito de las campañas comerciales, indicando con fundamentos científicos, a que clientes vender.

Se hizo a continuación un repaso general por el mundo del Data Mining, explicando todos los pasos del proceso que da origen a los modelos, resta ahora explicar, de que manera se obtuvo el modelo que permitió incrementar las tasas de éxito en la colocación de Plazos Fijos en forma significativa para el Banco.

7.1 Requisitos

Para aplicar Data Mining en la resolución de cierto problema, es necesario asegurar tres aspectos fundamentales de esta estructura: Software, Hardware y Datos.

7.1.1 Software

El Software utilizado elegido fue el SPSS Clementine.

La elección se realizó principalmente en base a la gran popularidad que tiene el programa entre los consumidores. Este software, además de ser el más popular, se ubica entre los que ofrecen mayor cantidad de funcionalidades.

7.1.2 Hardware

No fueron necesarias inversiones en computadoras. En función de los requisitos para el software elegido, se determinó que tanto la capacidad de procesamiento instalada como la memoria en disco eran suficientes para soportar la instalación y utilización del mismo.

7.1.3 Datos

Por razones de seguridad un Banco registra y realiza back-ups automáticos de todas las operaciones que realizan sus clientes (consumos con tarjeta de crédito, tasa del préstamo solicitado, cantidad de Plazos Fijos colocados en los últimos 5 años entre otros) y se guardan por un período no menor a 2 años.

Las variables sociodemográficas como edad, sexo y lugar de residencia son registradas al momento del alta de la cuenta del cliente.

7.2 Etapas de Data Mining

7.2.1 Selección gruesa de variables

Lo primero que se tiene al comienzo de un proyecto de Data Mining es una base de datos con una cantidad de variables desconocidas dispersas en tablas que tampoco se conocen.

La selección "gruesa" de variables consiste en determinar que tipo de variables se necesita recolectar en función del problema planteado.

Es necesario ser cuidadoso en esta primera selección de modo de no dejar afuera ninguna variable que pudiera ser de importancia para la creación del modelo. El resultado (output) de esta tarea debe ser un "set" de variables de número igual o inferior al original.

Se eligieron 102 variables (tabla 1) del total disponible en las bases del banco entre las cuales se encuentra la variable objetivo, "Tenencia de Plazo Fijo" (SI / NO). Por su naturaleza se las puede agrupar en: Identificación (1), Características (5), Sociodemográficas (4), Cantidades de Producto (22),

Tenencia de Productos (22), Productos y Consumos de los productos Bancarios (48).

Nombre del atributo	Descripción	Tipo
clinum	código de identificación del cliente	Identificación
fecha alta	fecha de ingreso	Características
operación usd	operación en dolares	
canal usd	canal de ingreso	
mes	mes	
origen	origen	
NSE	nivel socio economico (A, B, etc.)	
sexo	genero sexual	Sociodemográficas
region	region de residencia	
edad	edad	
cant pp	cantidad actual de prestamos personales	Cantidad
cant p acord	cantidad actual de prestamos pre acordados	
cant ph	cantidad actual de prestamos hipotecarios	
cant ref pp	cantidad actual de prestamos personales refinanciados	
cant CC	cantidad de cuentas corrientes	
cant CA	cantidad de cajas de ahorro	
cant DA	cantidad de debitos automaticos	
cant CJ	cantidad de conjuntos	
cant SV	cantidad de servicios	
cant SA	cantidad de sucursales asociadas	
cant MC	cantidad de Master Card	
cant MC TIT	cantidad de Master Card titulares	
cant MC ADI	cantidad de Master Card adicionales	
Cant SH	cantidad de seguros de hogar	
Cant SAU	cantidad de seguros de autos	
cant otros atm	cantidad de seguros ATM	
cant TD	cantidad de tarjetas de debito	
cant PMC	cantidad de productos/mantenimiento/cuentas	
cant ATM	cantidad de transacciones por ATM	
cant caja	cantidad de transacciones por caja	
cant IP	cantidad de transacciones por call center	
cant HB	cantidad de transacciones por home banking	

Tabla 1: Variables seleccionadas por filtro "grosso"

Flag PP	tenencia actual de prestamos personales	Tenencia (variables bancarias SI o NO)	
Flag P ACORD	tenencia actual de prestamos pre acordados		
Flag PH	tenencia actual de prestamos hipotecarios		
Flag ref PP	tenencia actual de prestamos personales refinanciados		
Flag CC	tenencia de cuentas corrientes		
flag ca	tenencia de cajas de ahorro		
flag DA	tenencia de debitos automaticos asociados a caja de ahorro		
flag cj	tenencia de conjuntos		
flag sa	tenencia de cuentas en sucursales asociadas		
flag sv	tenencia de servicios		
flag mc	tenencia de tarjetas master card		
flag mc tit	tenencia de tarjetas master card titular		
flag mc adi	tenencia de tarjetas master card adicionales		
flag sh	tenencia de seguros de hogar		
flag sau	tenencia de seguros automotor		
flag otros atm	tenencia de seguros atm		
flag td	tenencia de tarjetas de debito		
flag PMC	tenencia de productos/mantenimiento/cuentas		
flag ATM	tenencia de ATM		
flag caja	tenencia de caja		
flag IP	tenencia de call center		
flag HB	tenencia de home banking		
ca vigente	caja de ahorro vigente		PRODUCTOS Y CONSUMO
ca sum cred	cuenta crediticia		
ca sum deb	caja de ahorro con debito en cuenta		
sdo ca sum arp usd	saldo en caja de ahorro		
sdo ca prom arp	saldo promedio en caja de ahorro arp		
sdo ca prom usd	saldo promedio en caja de ahorro usd		
sdo ca prom eur	saldo promedio en caja de ahorro eur		
sdo cc prom arp	saldo promedio en cuenta corriente arp		
sdo cc prom usd	saldo promedio en cuenta corriente usd		
sdo cc prom eur	saldo promedio en cuenta corriente eur		
sdo cc fin de mes	saldo a fin de mes en cuenta corriente		
haberes	monto promedio de haberes		
jubilaciones	monto promedio de jubilación		
pf sum cap arp usd	plazo fijo suma de capital arp y usd		
pf cap prom arp	plazo fijo promedio de capital arp		
pf cap prom usd	plazo fijo promedio de capital usd		
pf cap prom eur	plazo fijo promedio de capital eur		
pmo per sdo vencer	prestamo personal saldo a vencer		
pmo per sdo vencido	prestamo personal saldo vencido		
pmo pre sdo vencer	prestamo prefinanciado saldo a vencer		
pmo pre sdo vencido	prestamo prefinanciado saldo vencido		
pmo re sdo vencer	prestamo refinanciado saldo a vencer		
pmo re sdo vencido	prestamo refinanciado saldo vencido		
segau	monto seguro automotor		
seghog	monto seguro hogar		
tc cons arp	consumo de tc en arp		
tc cons usd	consumo de tc en usd		
tc sum cons arp	suma consumo de tc en arp		
tc sum sal usd	suma saldo de tc en usd		
tc cant adic	consumo de tc adicional		
tc limite v	tarjeta visa limite		
tc limite m	tarjeta master limite		
tc finan v	tarjeta financiada visa		
tc finan m	tarjeta financiada master		
tc dnv v	tarjeta visa renovacion		
tc dnv m	tarjeta master renovacion		
tc compras v	tarjeta visa compras		
tc compras m	tarjeta master compras		
tc compras arp usd	tarjeta compras en arp y usd		
sdo cc prom arp	saldo en cc promedio arp		
td cant	cantidad detrajetas de debito		
td monto	monto por tarjeta de debito		
tc limite total	limite total en tarjetas de credito		
tc finan total	total financiado por tarjeta de credito		
tc dnv total	total renovacion de tarjetas de credito		
pp monto total	monto total de prestamos personales		
pr monto total	monto total de prestamos prefinanciados		
re monto total	monto total de prestamos refinanciados		

Tabla 1 (cont.): Variables seleccionadas por filtro "grueso"

7.2.2 Pre-Procesado de Datos

Cuanto mayor tamaño tiene una base de datos, mayores son las probabilidades de que existan errores en el registro de los mismos: "Ruidos", valores faltantes o valores "atípicos" son situaciones comunes que se encuentran en las bases, dado su gran tamaño y las diferentes fuentes que impactan registros en ellas.

Es posible encontrarse con casilleros en blanco o valores que no se corresponden con el sentido común, como por ejemplo, edades negativas, ausencia del nombre, números de documento que incluyen letras, entre otras. Los datos en esa forma no son aptos para ser procesados.

En un proyecto que involucra Data Mining, la preparación de los datos es el conjunto de pasos que mas tiempo abarca (ver gráfico 2). En general insume el 60% del tiempo total del proyecto.



Gráfico 2: Distribución del tiempo en etapas del Data Mining

Para realizar el pre-procesado de Datos, existen distintas técnicas que se aplican en forma automática o manual: *Limpieza, integración y normalización*.

La *limpieza* de datos se utiliza para remover el ruido, reemplazar valores faltantes y corregir inconsistencias en los datos. En la *integración* se reúnen datos provenientes de distintas bases y se los almacena en único lugar (almacén de datos). Por último, la *normalización* de los datos consiste en reducir la escala de valores de una variable manteniendo la relación de

magnitudes. Estas técnicas no son excluyentes entre sí sino que se complementan.

7.2.2.1 Limpieza de Datos

En el pre-lavado, se filtraron aquellas variables que por contener valores faltantes, ruido, o valores atípicos, no eran factibles de utilización en la generación del modelo.

Cabe aclarar que el caso de los Bancos es un caso especial ya que la correcta operación del negocio esta basada casi por completo en el flujo de información. Es por esto que la etapa de limpieza puede ser, como ocurrió en el presente trabajo, omitida. Esto no significa tomar datos de mala calidad sino, en vez de corregir los errores que pudiera haber, eliminar las variables y registros inviables mediante el prelavado.

7.2.2.2 Integración de Datos

En el Banco en estudio, se identifica a los clientes según un código alfanumérico de 6 caracteres único. Esta fue la variable que permitió la integración de todas las tablas.

7.2.2.3 Normalización de Datos

No se llevo a cabo una normalización de los Datos. Como se explicó anteriormente, no siempre se aplican indefectiblemente todos los pasos sino que en función de las necesidades se determina cuales realizar. En este caso las variables no mostraron diferencias significativas a priori en las magnitudes relativas como para esperar la que unas predominen sobre otras.

7.2.3 Selección Fina de Variables

7.2.3.1 Prelavado

Una vez realizada la integración y consolidación de todas aquellas variables elegidas en la selección gruesa en una única tabla, se procedió al proceso de prelavado. Mediante este proceso lo que se buscó es asegurar que las variables que iban a ser analizadas cumplieran con determinadas condiciones mínimas de aceptación, tal que los resultados obtenidos en forma posterior tuviesen significancia estadística.

En este paso, se deben definir los criterios de filtro para las variables y registros tal que se eliminen aquellas y aquellos que no aportan información útil. Las opciones de filtro se basan en los valores observados de la variable en análisis, sin contemplar la eficacia predictiva de la misma.

A continuación se muestra como se realiza el prelavado utilizando el software.

Se pueden elegir entre 0 y 5 criterios de filtro o cribado (ilustración 7). Es posible modificar los valores del umbral que el programa recomienda por default.

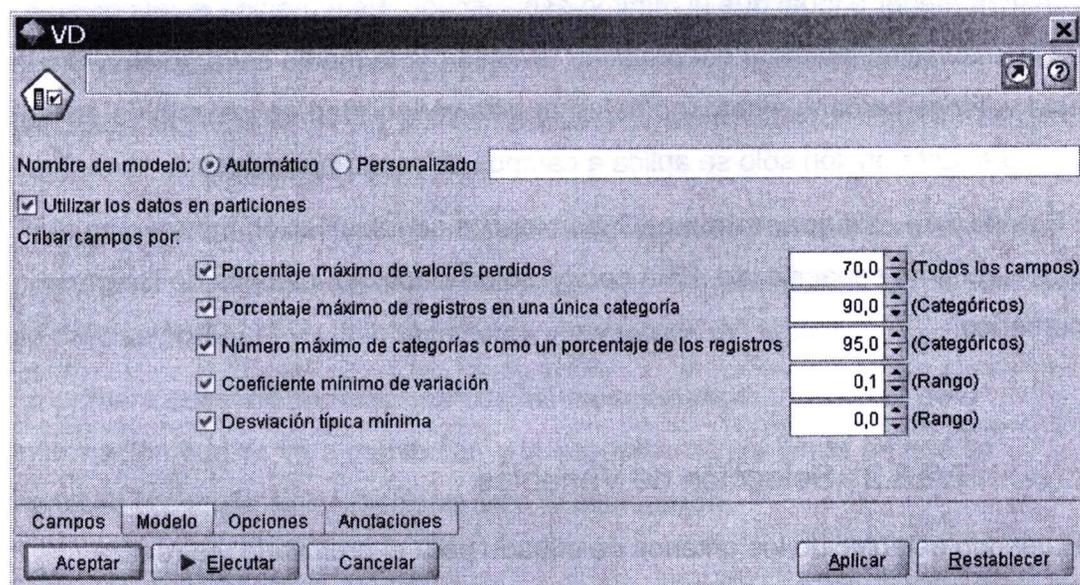


Ilustración 7: Selección de criterios y valores umbral en Prelavado

Descripción de los criterios:

- **Porcentaje máximo de valores perdidos:** Criba campos con demasiados valores perdidos, expresados como un porcentaje del número total de registros. Los campos con un alto porcentaje de valores perdidos proporcionan poca información predictiva.
- **Porcentaje máximo de registros en una categoría única:** Criba campos con demasiados registros dentro de la misma categoría en relación con el número total de registros. Por ejemplo, si el 95% de los clientes de la base de datos conduce el mismo tipo de coche, no sería útil incluir esta información para distinguir a un cliente de otro. Cualquier campo que exceda el máximo especificado se criba. Esta opción sólo se aplica a campos categóricos.
- **Número máximo de categorías como un porcentaje de registros:** Criba campos con demasiadas categorías en relación con el número total de registros. Si un porcentaje elevado de las categorías contiene sólo un único caso, puede que el campo sea de uso limitado. Por ejemplo, si cada cliente lleva un sombrero diferente, será improbable que esta información sirva a la hora de modelar patrones de comportamiento. Esta opción sólo se aplica a campos categóricos.
- **Coefficiente mínimo de variación:** Criba campos con un coeficiente de varianza menor o igual que el mínimo especificado. Esta medida es el cociente de la desviación estándar del predictor dividida por la media del predictor. Si este valor es cercano a cero, no habrá mucha variabilidad en los valores de la variable. Esta opción sólo se aplica a campos de rango numérico.
- **Desviación estándar mínima:** Criba campos con desviación menor o igual que el mínimo especificado. Esta opción sólo se aplica a campos de rango numérico.

7.2.3.2 Selección de Variables

Habiendo establecido los criterios de cribado para el prelavado, se realiza sobre los registros y variables que no fueron excluidas, el filtrado de estas últimas pero ahora en función de su calidad como predictoras. Para hacerlo se

utilizó un modelo *tipo filtro*, es decir, se asigna a cada variable valores de importancia predictiva sin importar como interactúan con el algoritmo de predicción que se utilizará en forma posterior para la creación del modelo.

7.2.4 Transformación

El primer paso para la preparación de datos es conocer el problema a resolver, para lo cual se debió incluir como actividad preliminar la comprensión del dominio: el propósito es asegurar el entendimiento del negocio y objetivos del proyecto, o al menos hacia qué objetivo queremos llegar, sin esto resulta imposible conocer los datos que se deben extraer. Por otra parte se necesita conocer la forma en que se debe presentar la información al modelo seleccionado para la explotación de datos, con estas dos precisiones se está en condiciones para comenzar a recolectar la información y trabajar con ella. Cuando se trabaja en explotación de datos, se utilizan datos que representan hechos de la vida real y que se preparan para que puedan ser utilizados en el modelo.

La preparación de los mismos no es un proceso automático, por lo cual es necesario aplicar propio conocimiento para generar el conjunto de datos necesario para poder desarrollar un modelo de explotación. Por lo antes dicho, el principal objetivo de la preparación de datos es tomar información, manipularla, transformarla y presentarla para que pueda ser procesada por un modelo de data mining.

Para conocer que transformaciones se deben realizar y en que forma hay que presentarla, se deben responder dos preguntas fundamentales: ¿Qué solución se precisa obtener? y ¿Qué técnica de explotación se utilizará?

La primera cuestión se relaciona con las características y cantidad de información que se va a manipular, y la segunda, con la forma en que se necesita presentar la información para la explotación.

7.2.5 Creación de Modelos

Una vez seleccionadas las variables mas relevantes para la predicción, se procede con la generación del modelo predictivo.

Los algoritmos probados fueron:

- Redes Neuronales:

Las redes neuronales son modelos simples del funcionamiento del sistema nervioso. Las unidades básicas son las neuronas, que generalmente se organizan en capas, como se muestra en la siguiente ilustración:

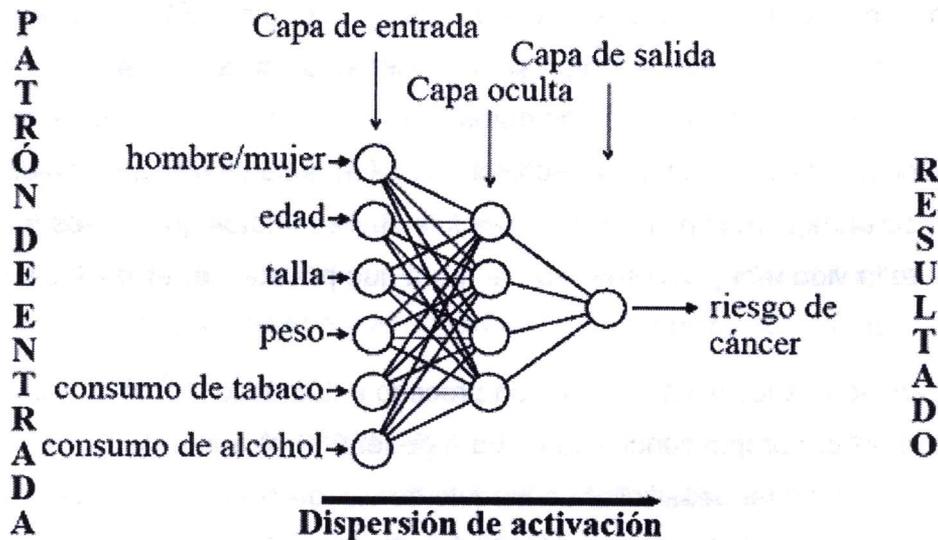


Ilustración 8: Estructura de una red neuronal

Una red neuronal, a menudo denominada perceptrón multicapa, es básicamente un modelo simplificado del modo en que el cerebro humano procesa la información. Funciona combinando en forma simultánea un número elevado de unidades simples de procesamiento interconectadas que parecen versiones abstractas de neuronas.

Las unidades de procesamiento se organizan en capas. Existen, generalmente, tres capas en una red neuronal: una capa de entrada, con unidades que representan los campos de entrada; una o varias capas ocultas; y una capa de salida, con unidades que representan los campos de salida. Las unidades se unen con fuerzas de conexión variables o ponderaciones. Los datos de entrada

se presentan en la primera capa, y los valores se propagan desde cada neurona hasta cada neurona de la capa siguiente. Al final, se envía un resultado desde la capa de salida.

La red aprende examinando los registros individuales, generando un pronóstico para cada registro y realizando ajustes a las ponderaciones cuando realiza un pronóstico incorrecto. Este proceso se repite varias veces y la red sigue mejorando sus pronósticos hasta haber alcanzado uno o varios criterios de parada.

Al principio, todas las ponderaciones son aleatorias y las respuestas que resultan de la red son, posiblemente, insólitas. La red aprende a través del entrenamiento. Continuamente se presentan a la red ejemplos para los que se conoce el resultado, y las respuestas que proporciona se comparan con los resultados conocidos. La información procedente de esta comparación se pasa hacia atrás a través de la red, cambiando las ponderaciones gradualmente. A medida que progresa el entrenamiento, la red se va haciendo cada vez más precisa en la replicación de resultados conocidos. Una vez entrenada, la red se puede aplicar a casos futuros en los que se desconoce el resultado.

- C5:

El algoritmo C5.0 genera un árbol de decisión o un conjunto de reglas. El modelo divide la muestra basándose en el campo que ofrece la máxima ganancia de información en cada nivel. El campo objetivo debe ser categórico y se permiten varias divisiones en más de dos subgrupos.

- CHAID:

El algoritmo CHAID genera árboles de decisión utilizando estadísticos de chi-cuadrado para identificar las divisiones óptimas. A diferencia de otros algoritmos, CHAID puede generar árboles no binarios, lo que significa que algunas divisiones generarán más de dos ramas. Los campos objetivo y predictor pueden ser de rango o categóricos.

- Combinación (C5 + Red Neuronal + CHAID)

A continuación se procedió con la evaluación de cada modelo en forma individual y por último la combinación de ellos.

Para hacerlo se determinó al azar una muestra de comprobación sobre la cual se corrieron los 4 modelos y se observaron los resultados.

7.2.6 Combinación de modelos parciales para obtener el modelo final

Se combinaron 3 algoritmos para obtener pronósticos más precisos de los que pueden conseguirse de los modelos individuales. Al combinar predicciones de varios modelos, pueden evitarse las limitaciones en modelos individuales que dan como resultado una precisión global superior. Los modelos combinados de esta forma suelen ejecutarse tan bien como el mejor de los modelos individuales y, en ocasiones, mejor.

El método utilizado es la votación, que funciona cuadrando el número de veces que cada posible valor pronosticado se elige y seleccionando el valor con mayor valor total. Por ejemplo, si tres de los cinco modelos pronostican sí y los otras dos pronostican no, se gana por 3 a 2. A dicho registro se le asigna la mayor probabilidad individual que hayan arrojado los 3 modelos que votaron sí.

Cabe destacar que no es objetivo de este trabajo ahondar en las técnicas de Data Mining.

8 Implementación y Análisis de Resultados

8.1 Prueba del Modelo en caso Real

Una vez obtenido el modelo se lo puso a prueba.

La forma de determinar el impacto de la implementación del modelo para la selección de clientes es utilizándolo en una campaña real.

Se seleccionaron 2 grupos de clientes:

El primero, el grupo testigo, se seleccionó según la vieja estrategia.

El segundo se eligió alimentando el modelo con la tabla de clientes y luego de obtener las probabilidades de aceptación de cada cliente, se eligieron aquellos más propensos.

Luego de realizar las ofertas comerciales a todos los clientes, se midió la tasa de éxito de ambos grupos, es decir, el porcentaje de ventas logrado sobre el total de clientes contactados en cada estrategia. Los resultados se muestran en la siguiente tabla:

Campaña Plazo Fijo					
SIN MODELO (testigo)			CON MODELO		
Contactados	Vendidos	Tasa de Éxito	Contactados	Vendidos	Tasa de Éxito
2000	350	18%	2000	720	36%

Tabla 2: Comparación entre ambas estrategias de selección de clientes

8.2 Impacto económico logrado

8.2.1 Ingresos Generados

Según la información obtenida, en el 2008 se dieron de alta por campañas comerciales un total de **2000 Plazos Fijos**⁷.

⁷ Informe de Performance 2008 - Campañas Comerciales Banco en estudio

A continuación se muestra la tabla⁸ que contiene los parámetros necesarios para calcular el rendimiento esperado de un plazo fijo (la ganancia resultante por la venta):

Tasa que se paga al cliente (TNA)	11,70%	
Plazo promedio de colocación (meses)	4	
Tasa cobrada a la mesa de dinero (TNA)	15,15%	
Spread anual	3,45%	(tasa cobrada - tasa pagada)
Spread mensual capitalizado	0,2831%	
Monto promedio colocado en plazos fijos (\$)	35000	
Beneficio por PF vendido (\$)	403	(Monto promedio * Spread mensual * Plazo promedio de colocación)

Tabla 3: Cálculo de Beneficio por PF vendido

Esto significa que cada Plazo Fijo adicional que se vende representa para el Banco un ingreso de \$403.

Para calcular el volumen esperado de ventas en 2010, se supondrá que durante 2010 la cantidad de clientes contactados, el contexto, los canales utilizados y el atractivo del producto son similares a los del 2008.

Como se asumen igual cantidad de contactos en el 2010, las ventas de plazos fijos se calculan como:

Volumen de Ventas 2010 (con modelo) = $2000 * 36\% / 18\% = 4000$ Plazos Fijos
--

Comparando ambos escenarios, sin modelo y con modelo se observa que, gracias a la implementación del modelo, se logra un incremento en las ganancias del Banco de \$ 806.000 anuales.

	Volumen	Ganancia Generada
Plazos Fijos dados de alta 2008 sin modelo	2000	\$ 806.000
Plazos Fijos dados de alta 2010 con modelo	4000	\$ 1.612.000
Incremento en la Ganancia del Banco	2000	\$ 806.000

Tabla 4: Cálculo del incremento en la Ganancia del Banco por la implementación de Modelo de Respuesta

⁸ Datos de Inteligencia Comercial - Banco en estudio

8.2.2 Costos incurridos

Los costos adicionales en los que se incurrió para lograr la solución tienen tres componentes: Licencia del Software, Hardware y Mano de Obra. A continuación se detalla cada uno de ellos.

8.2.2.1 Software

La licencia para la adquisición del software tiene un valor de \$60.000. Calculando una vida útil de 5 años y realizando la división simple entre el monto pagado y los años de utilización se concluye que:

$$\text{Costo anual del software} = \$60.000 / 5 \text{ años} = \mathbf{\$12.000}$$

Se asume excluir el costo del software dado que el mismo se amortiza aplicándolo en infinidad de campañas a lo largo de la vida del Banco. Suponiendo que el período de amortización fuese un año, y se tiene previsto realizar un total de 20 campañas ese año, y que el costo a imputar a una campaña se reparte equitativamente, tenemos que solamente 1/20 del costo del software se debe imputar a la campaña; o sea \$ 2.500.

8.2.2.2 Hardware

Si bien no fue necesaria la instalación de equipos adicionales, se considera que en algún momento será necesaria la compra de una nueva computadora para reemplazar la asignada a los proyectos de Data Mining. Se supone una vida útil de 3 años.

A precios actuales, un equipo como el descrito según los requerimientos del Software tiene un costo de:

$$\text{Costo anual Hardware (PC)} = \$ 4.000 / 3 \text{ años} = \mathbf{\$ 1.334}$$

8.2.2.3 Mano de Obra

El tiempo de dedicación necesario para lograr el modelo predictivo y luego acompañar en todas las etapas de implementación, control, evaluación y

actualización es de **4 meses efectivos**. Actualmente el sueldo percibido por este profesional es de \$10.000 pero al Banco le cuesta **\$12.000** debido a las cargas sociales. Dado que trabaja solo 4 meses al año abocado a los modelos de Plazos Fijos, se tiene que:

$$\text{Costo Anual Mano de Obra} = \$12.000 / \text{mes} * 4 \text{ meses} = \$ 48.000$$

8.2.3 Impacto económico

Como se explicó anteriormente y dado los ingresos potenciales de nuevas ventas por plazo fijo, se suponen que los ingresos adicionales generados serán de \$ 806.000/año.

También se contemplan los costos adicionales incurridos (\$12.000 (Software) + \$1.334 (Hardware) + \$48.000 (Mano de Obra)) que son de \$ 61.334, esto hace un beneficio directo de \$744.667 por año.

GANANCIA ANUAL GENERADA	\$ 806.000
<i>Software</i>	<i>(\$ 12.000)</i>
<i>Hardware</i>	<i>(\$ 1.334)</i>
<i>Mano de Obra</i>	<i>(\$ 48.000)</i>
SUBTOTAL COSTOS INCURRIDOS	<u>- \$ 61.334</u>
BENEFICIO ANUAL LOGRADO	\$ 744.666

9 Conclusiones y futuras líneas de investigación

Se probó que la tasa de éxito utilizando el modelo supera ampliamente la obtenida mediante la vieja estrategia de supuestos y selección aleatoria.

La tasa de éxito se multiplicó al doble, logrando aumentar las ventas en igual proporción, quedando demostrada la eficiencia de esta herramienta mediante el aumento en las ventas de Plazos Fijos, como así también el poder del Data Mining, el cual puede ser aplicado a muchos otros casos de negocios dentro de la Empresa.

Dado que la metodología a seguir para lograr el modelo es similar para cada proyecto de Data Mining, el procedimiento descrito en el presente trabajo puede ser utilizado para la resolución de nuevos proyectos a encarar.

Queda demostrado también que la "experiencia" de un individuo en cierto campo puede ser reemplazada, y con mejores resultados, por datos históricos, los cuales mediante un análisis apropiado, aportan conocimiento con validez estadística. Un análisis científico riguroso como el aportado por esta técnica logra desentrañar relaciones ocultas hasta para el ojo de los "expertos" y permite tomar decisiones más acertadas que impactan directamente en los ingresos percibidos por la empresa.

Los puntos a tener en cuenta al momento de implementar esta solución son los cambios de propensión al consumo, que como mitigante se deberán ajustar los algoritmos para analizar cada caso en particular o en su defecto trabajar en aumentar la necesidad de consumo. Por otro lado también se deberá estar atento a los cambios de escenarios que se podrían dar en un entorno macro económico y político altamente sensible.

Como futuras líneas de investigación se desprende la inquietud de analizar el impacto de dichos cambios de escenarios haciendo desarrollos de sensibilidad para dar más sustento a este tipo de soluciones.

10 Bibliografía

Libros

- Berson A. et al., 1999.** Building Data Mining Applications for CRM
Berry and Linoff, 2008. Data Mining Techniques for marketing, sales and customer relationship
Freeman J. A./ Skapura D.M., 1993. Redes neuronales: Algoritmos, aplicaciones y técnicas de programación
Kass, G.V., 1980. Chi-squared Automatic Interaction and Detection (CHAID)
Quinlan, J.R., 1986. Induction of Decision Trees

Papers

- C. Olivia Rud. Data Warehousing for Data Mining: A Case Study.** Páginas 119-125
Chapman et al., 1999. CRISP-DM 1.0 Step-by-step data mining guide

Note on Data Mining and BI in Banking Sector

<http://www.iimahd.ernet.in/publications/data/Note%20on%20Data%20Mining%20&%20BI%20in%20Banking%20Sector.pdf>

- Ramin Mikaili et al., 2000.** Data Mining: An implementation referente guide
Xindong Wu et al., 2008. Top 10 algorithms in data mining

Un método de transformacion de datos orientado al uso de explotación de información. H. Merlino, P. Britos, J. Ierache, E. Diez y R. García-Martínez
<http://www.itba.edu.ar/archivos/secciones/62CACIC-2005-Metodo-de-Transformacion-de-Datos.pdf>

Noticias

Biotech companies testify on data mining

<http://www.milforddailynews.com/news/x148103593/Biotech-companies-testify-on-data-mining>

Business intelligence o el fin de las empresas "tontas"

<http://tecnologia.infobaeprofesional.com/notas/68301-Business-intelligence-o-el-fin-de-las-empresas-tontas.html?cookie>

Los beneficios de la inteligencia de negocios

<http://tecnologia.infobaeprofesional.com/notas/51865-Los-beneficios-de-la-inteligencia-de-negocios.html>

La inteligencia de datos no está madura en la Argentina

<http://tecnologia.infobaeprofesional.com/notas/43472-La-inteligencia-de-datos-no-esta-madura-en-la-Argentina.html>

¿Para qué sirve el "business intelligence"?

<http://www.infobaeprofesional.com/notas/80821-Para-que-sirve-el-business-intelligence.html&cookie>

Research and Markets: Applied Data Mining for Business and Industry, 2nd Edition

<http://www.tradingmarkets.com/.site/news/Stock%20News/2270751/>