
PREDICCIÓN DE FUGAS DE CLIENTES PARA UNA INSTITUCIÓN FINANCIERA MEDIANTE SUPPORT VECTOR MACHINES

JAIME MIRANDA*
PABLO REY*
RICHARD WEBER*

Resumen

La fuga de clientes es un fenómeno que atañe a la gran mayoría de las instituciones financieras, siendo también un tema de intensivo estudio científico en los últimos años. En el presente trabajo se presenta un modelo predictivo desarrollado para identificar los clientes con tendencias a la fuga en un banco. De esta forma es posible hacer más efectivas las políticas comerciales de retención, ser más eficientes en la asignación de los recursos al focalizar a los ejecutivos y mejorar las relaciones con los clientes al detectar los principales focos de deficiencias del servicio. Estos beneficios generan una mayor rentabilidad al aumentar tanto el número de transacciones como la permanencia del cliente. El modelo presentado se basa en la técnica de Support Vector Machines. La comparación con un modelo de predicción basado en redes neuronales muestra la superioridad del modelo propuesto en el caso considerado.

Palabras Clave: Retención de clientes; Support Vector Machines.

1. Introducción

La cartera de clientes es uno de los activos más importantes para una institución financiera, ya que está estrechamente relacionada con las utilidades del negocio. Dos actividades comerciales tienen como objetivo mantener y mejorar dicha cartera: la captación de clientes nuevos y la retención de clientes existentes.

*Departamento de Ingeniería Industrial, Universidad de Chile

La *captación de clientes* apunta a aumentar el número de clientes de la cartera a través de la definición e incorporación de nuevos segmentos objetivos. Esta captación se realiza principalmente a través de elaboradas estrategias de publicidad, alta inversión en fuerza de ventas y la generación de ofertas focalizadas. La *retención de clientes* consiste en la identificación de los clientes con mayores tendencias a la fuga y en la determinación de las estrategias o procedimientos que aumenten el grado de fidelización y bajen los índices de fuga en la cartera.

Existen dos tipos de fuga: las fugas voluntarias y las fugas no voluntarias. Las *fugas voluntarias* se asocian a la desafiliación del cliente por iniciativa propia, sin injerencia directa por parte de la institución. A diferencia del caso anterior, las *fugas no voluntarias* son desafiliaciones en donde el banco es responsable directo del término de los acuerdos contractuales, donde el cliente no posee ninguna injerencia. Este tipo de cierre se gatilla principalmente por acciones delictuales o por mala utilización de los productos. Un ejemplo puede ser efectuar algún tipo de fraude financiero con cheques del banco o la clonación de tarjetas de crédito. En el presente trabajo estudiaremos las fugas voluntarias.

La sección 2 discute la importancia de la retención de clientes y el estado del arte en la literatura respecto del tema. Las secciones siguientes detallan la aplicación de una metodología a un caso particular en un banco chileno. En la sección 3 se discuten las variables utilizadas en el estudio. La sección 4 presenta el preprocesamiento y las transformaciones aplicadas a estas variables. En la sección 5 se desarrolla el modelo matemático para la predicción de fuga. La sección 6 presenta los resultados obtenidos comparándolos con los de un modelo alternativo. En la sección 7 se concluye el trabajo presentado y se exponen líneas de trabajo futuro.

2. ¿Por qué es Importante Retener a un Cliente?

Estudios recientes [1, 9, 13] han mostrado los beneficios aportados por la retención de clientes en instituciones bancarias. Al tener una cartera con un mayor número de clientes, se realizan en promedio un mayor número de transacciones, aumentando las utilidades de la institución. Además, si se logra una mayor permanencia de un cliente en la institución, se obtienen los beneficios asociados a la disminución de los costos operacionales, las referencias y al incremento en las transacciones.

La figura 1 muestra estos efectos y el beneficio monetario generado por un cliente a través del tiempo [16]. Sólo en el primer año un cliente no reporta beneficio a la institución. Los costos de este periodo corresponden a la incor-

poración e inicialización del cliente y a la inversión en las distintas campañas de captación.

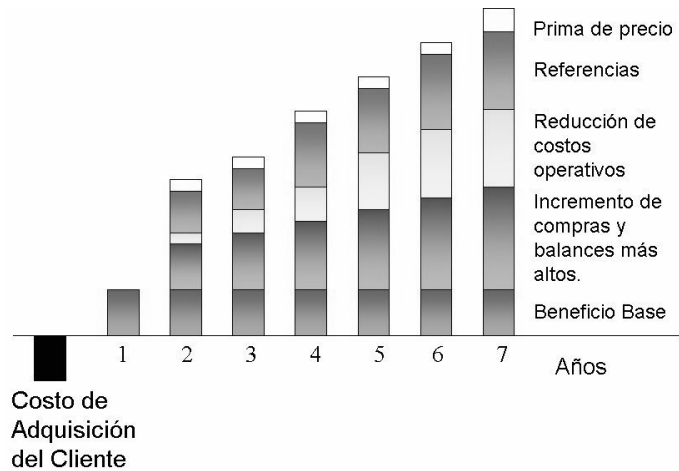


Figura 1: Beneficio generado por un cliente a lo largo del tiempo que permanece dentro de la institución [16].

La predicción de fuga es un elemento importante para la retención de clientes. Tanto la identificación de los clientes con tendencias a fuga como la determinación de su rentabilidad futura permiten focalizar los esfuerzos de retención en los clientes más apropiados [9, 13]. La figura 2 ilustra el ciclo de retención basado en una adecuada predicción de fuga.

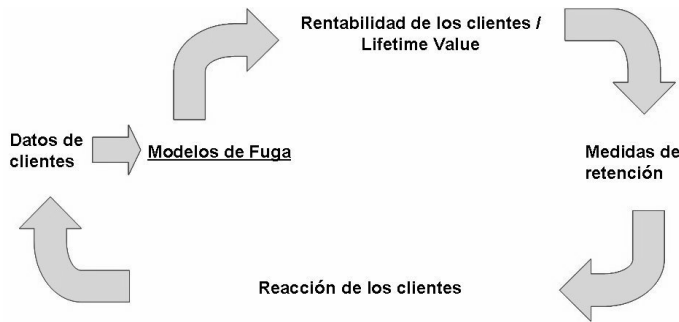


Figura 2: Ciclo de retención de clientes.

Un estudio realizado en el banco ECB (European Central Bank) [18] muestra los beneficios generados al disminuir el porcentaje de clientes fugados anualmente y cómo se ven afectados los flujos futuros de la institución durante un periodo de 25 años. En dicho estudio se definen la *tasa de retención* como la relación entre el número de clientes que permanecen en la institución y la cartera total de clientes.

$$\text{Tasa de retención} = \left(1 - \frac{\text{número de clientes fugados al año}}{\text{número total de clientes en cartera}}\right) \cdot 100\% \quad (1)$$

La figura 3 exhibe el beneficio acumulado generado por la cartera respecto de distintas tasas de retención. Se observa que a medida que aumenta la tasa de retención, el beneficio también aumenta. La situación ideal es la representada por la línea correspondiente a una tasa de retención igual al 100 %. Esto significaría que se retiene a la totalidad de los clientes cada año. Esta situación es prácticamente imposible debido a que existen numerosas circunstancias que el banco no maneja, por ejemplo, la muerte de un cliente o el cambio de su residencia al extranjero.

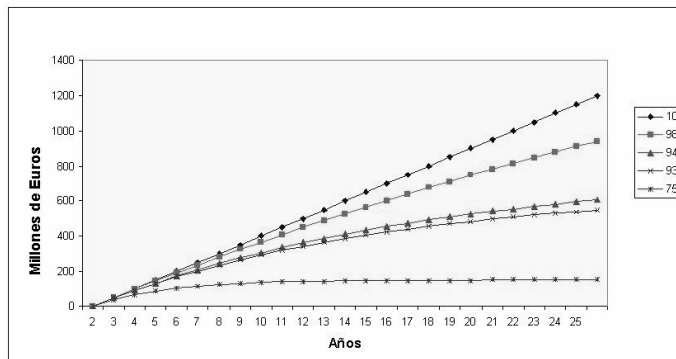


Figura 3: Beneficio acumulado respecto de la tasa de retención [18].

Se observa que con un aumento de la tasa de retención de un 1% (de un 93% a un 94%), respecto de una cartera de 1.000.000 de clientes, las utilidades generadas por la retención de este segmento crecen en aproximadamente 30 millones de euros durante el periodo de evaluación [18].

Cabe destacar que no es viable remediar la fuga de clientes con una mayor captación de clientes nuevos. Por un lado, los clientes nuevos son potencialmente riesgosos para la empresa. Por otro lado, se sabe que captar un cliente nuevo es entre 5 y 6 veces más costoso que retener a uno antiguo [3, 15].

En resumen, las actividades para retener clientes generan una serie de beneficios para una institución financiera que justifican el desarrollo de modelos sofisticados para la predicción de fuga.

3. Descripción de las Variables Utilizadas

En varios estudios para instituciones financieras se han identificado las variables más relevantes para la predicción de fuga de clientes. Para el presente trabajo, se comprobó la utilidad de estas variables usando métodos estadísticos, por un lado, y la experiencia de los expertos, por otro.

Finalmente se generó una agrupación de las variables, asignándolas a uno de los siguientes cuatro grupos:

1. Variables de Comportamiento Bancario.
2. Variables Sociodemográficas.
3. Variables de Percepción de la Calidad del Servicio.
4. Variables de Entorno.

A continuación se describe cada uno de estos grupos para la aplicación presentada en este artículo.

La muestra utilizada corresponde a la cartera completa de clientes que poseía la institución entre el 1 de enero y el 31 de diciembre del año 2002. Esta muestra incluye todos los clientes vigentes durante el periodo y todos los clientes que se fugaron durante el mismo periodo, excluyendo las fugas involuntarias.

3.1. Variables de Comportamiento Bancario

Entre las variables que capturan el comportamiento bancario de los clientes se pueden mencionar los saldos mensuales promedios en la cuenta corriente, el número de transacciones mensuales por los distintos canales de comunicación, la tenencia de productos y servicios y las variaciones en su utilización. El objetivo principal al utilizar este conjunto de variables es identificar cambios en los patrones de comportamiento transaccional de los clientes.

A modo de ejemplo, la figura 4 muestra el número de transacciones realizadas mensualmente por los distintos canales por los que opera la institución: caja, portal Web, call center, RedBanc y RedCompra. Se observa que el principal canal transaccional que utilizan los clientes es el portal Web (47%), seguido por las transacciones por call center (30%).

Se ha observado [14] que mientras más transacciones realiza un cliente, menor es su probabilidad de fuga.

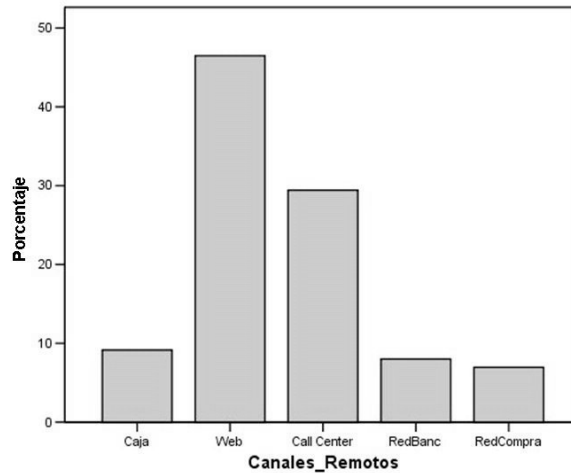


Figura 4: Distribución del número de transacciones por canal.

3.2. Variables Sociodemográficas

Este conjunto de variables corresponde a caracterizaciones relacionadas con aspectos personales de los clientes, por ejemplo, edad, renta, estado civil y nivel educacional.

En [12] se considera a la edad como variable discriminadora. En ese estudio se concluye que las personas de mayor edad son más estables en sus preferencias, y por ende, tienen menores tendencias a cambiarse a otra institución financiera. En caso contrario, se observa que las personas jóvenes son más inestables en sus preferencias aumentando su tendencia a cambiar de institución financiera. La figura 5 muestra la distribución de la edad de los clientes de la cartera total.

Cerca del 60% de la cartera tiene una edad entre los 25 y 35 años, lo que implica que la institución posee un segmento significativo de clientes jóvenes.

3.3. Variables de Percepción de la Calidad del Servicio

Existen numerosos trabajos que relacionan la calidad del servicio con los patrones de fuga [1, 10, 11] concluyendo que una mala percepción del servicio aumenta las tendencias a fugarse.

Una variable de este grupo es el número de reclamos realizados al centro de atención en un mes por el cliente. Mensualmente se realizan, en promedio, 350 reclamos. En la figura 6 se exhibe la distribución del número de reclamos realizados por un cliente en un mes particular. En esta figura sólo se representa a los clientes que realizaron reclamos, que constituyen un 1,5% de la cartera.

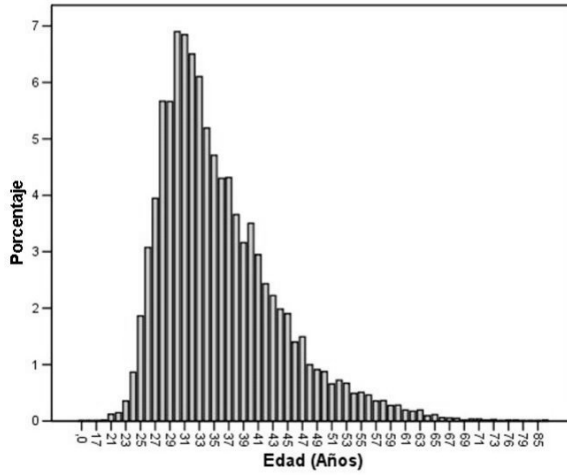


Figura 5: Distribución de la variable edad.

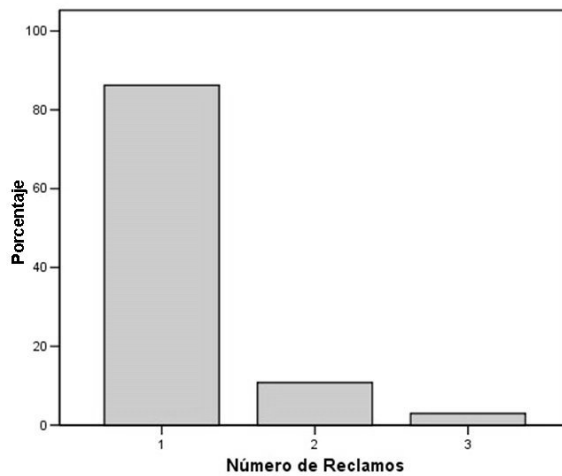


Figura 6: Distribución del número de reclamos por cliente realizados en forma mensual.

Con un análisis más detallado se pudo observar que más del 75% de los clientes que cerraban sus productos habían realizado por lo menos un reclamo al centro de atención de clientes (ver figura 7).

Esta relación entre el número de reclamos y el cierre de la cuenta corriente motiva la inclusión de esta variable para la identificación del patrón de fuga.

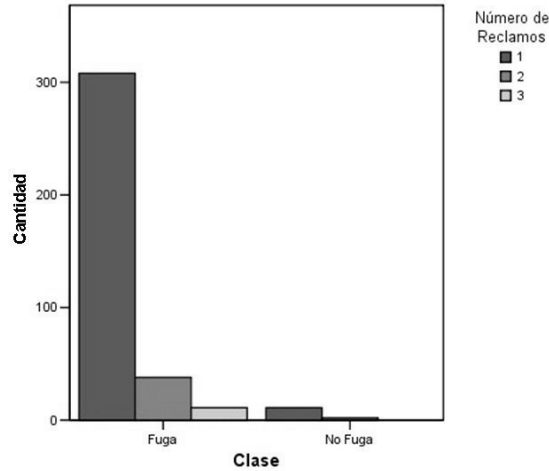


Figura 7: Relación entre el número de reclamos y las fugas mensuales.

3.4. Variables de Entorno

Estas variables caracterizan los efectos ocasionados por alguna perturbación dentro del mercado financiero capturando si el cliente sufrió un hito o acontecimiento especial durante su ciclo de vida dentro de la institución. Un ejemplo de estas variables es la *antigüedad* del cliente, la que corresponde a la cantidad de meses que lleva dentro de la institución. La figura 8 muestra la distribución de esta variable respecto a la cartera total.

En [18] se concluye que clientes con una mayor antigüedad tienen menor tendencia a cambiarse.

4. Preprocesamiento y Transformación de las Variables

La forma en que los datos aparecen originalmente en una base de datos muchas veces no representa de la mejor manera la información disponible. En general, primero hay que preprocesar los datos para corregir posibles errores e inconsistencias dentro de la base. Transformaciones posteriores pueden enriquecer aún más la información que proveen estos datos.

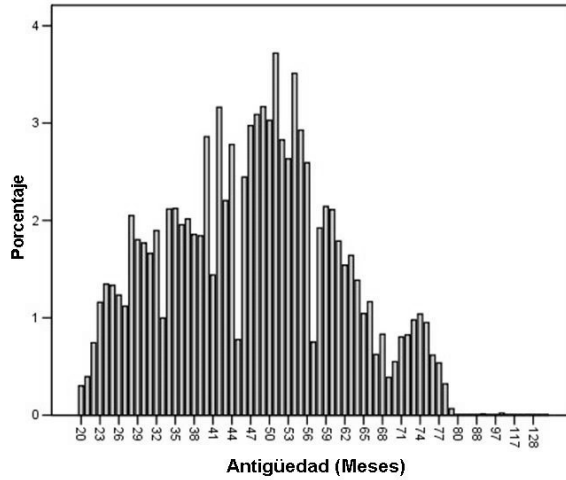


Figura 8: Distribución de la variable antigüedad.

4.1. Etapa de Preprocesamiento

El objetivo del preprocesamiento en este trabajo fue resolver el problema relacionado a los valores faltantes y fuera de rango dentro de la base de datos. Se entiende por *valores faltantes* a la ausencia del valor de cierta variable que describe al cliente y por *valores fuera de rango* a valores que se escapan de los rangos permitidos para una variable. Ejemplos de este último grupo son las edades mayores de 200 años o las rentas negativas. Entre las soluciones implementadas están:

- Eliminación de registros;
- Llenado con promedios y modas;
- Utilización de modelos predictivos.

La determinación de los rangos de las variables se realizó en forma conjunta con los expertos del negocio.

El resultado de esta etapa es una base de datos limpia y sin inconsistencias.

4.2. Etapa de Transformación

La base de datos utilizada no fue creada inicialmente con la visión de construir modelos matemáticos de predicción. Debido a esto, no todas las variables seleccionadas estaban en condiciones de ser utilizadas en su estado original. Por esta razón, fue necesario utilizar transformaciones de las variables originales para enriquecer la información contenida y para hacerlas compatibles con

las técnicas de modelamiento. Entre las principales transformaciones utilizadas podemos mencionar:

1. La transformación de las variables texto (categorías) a números. En el cuadro 1, a modo de ejemplo, se muestra la transformación utilizada para la variable *nivel educacional*.
2. La transformación de la variable *fecha de nacimiento* a *edad* (en años).
3. El escalamiento y estandarización de las variables numéricas dentro de un rango específico, por ejemplo en el rango $[0,1]$. Esta transformación apunta a tener las distintas variables en escalas comparables, lo que mejora la convergencia de los algoritmos de aprendizaje [4].
4. La generación de nuevas variables a partir del conjunto de variables inicial. Un ejemplo de este grupo pueden ser las variaciones porcentuales entre los saldos de meses consecutivos.

Nivel Educacional	
Categoría	Clase
Educ. Universitaria	1
Educ. Técnica	2
Educ. Media	3
Estudiante Universitario	4

Cuadro 1: Ejemplo de transformación para una variable.

El resultado de esta etapa es una base de datos transformada a valores numéricos, lista para los análisis estadísticos y para la construcción de los modelos matemáticos de minería de datos.

5. Construcción del modelo y enfoque de solución

El problema planteado se enfrentó con un enfoque de clasificación binaria [7]. Este tipo de procedimiento se basa en la determinación de una *función clasificadora* que permite asignar a cada objeto a una de las dos clases definidas *a priori*. En nuestro caso, cada cliente será asignado a una de las clases “fuga” o “no fuga”.

La construcción del modelo se lleva a cabo en dos etapas: entrenamiento y test. Para cada una de las etapas se considera un subconjunto del total de los objetos (clientes) a clasificar. Estos subconjuntos de objetos forman una partición del conjunto total de objetos y son llamados, *conjunto de entrenamiento* y *conjunto de test*, respectivamente.

En la etapa de entrenamiento se estima la mejor función clasificadora considerando algún criterio (por ejemplo, el error de clasificación) en el conjunto de entrenamiento. En la etapa de test se valida la efectividad del modelo respecto de objetos no utilizados en el entrenamiento. Para esto, se utiliza el modelo obtenido para clasificar los elementos del conjunto de test. El modelo asigna cada objeto a una de las clases definidas, la que llamaremos la “clase generada” del objeto en contraposición a la “clase real” que es la clase a la que el objeto efectivamente pertenece. Considerando los objetos “mal clasificados” (aquellos cuya clase generada es diferente a su clase real) se estima un error de clasificación. Dependiendo de este error, se revisa el modelo propuesto.

Existen diversas técnicas que tratan el problema de clasificación binaria. Entre estas podemos mencionar: las redes neuronales artificiales, los árboles de decisión y los Support Vector Machines (SVM) [7]. Para este trabajo, motivados por la efectividad y robustez en problemas de clasificación reportada en la literatura sobre este tipo de métodos (ver por ejemplo, [4, 6, 7]) se seleccionaron los SVM.

A continuación se describe brevemente esta técnica y se muestra la forma en que se obtiene la función de separación.

5.1. Support Vector Machines

La técnica de Support Vector Machines fue propuesta por Vapnik [19, 20]. Esta técnica se basa en encontrar un hiperplano de separación que divida el espacio de entrada en dos regiones. Cada una de estas regiones corresponderá a una de las clases definidas.

Por un lado, a hiperplanos que están más alejados de las fronteras de las clases de objetos corresponden mayores márgenes de separación. Por otro, hiperplanos que aciertan más en la asignación de objetos a las clases a las que

efectivamente pertenecen, tienen un menor error de clasificación. Por lo tanto, un hiperplano de separación ideal debe maximizar el margen de separación y minimizar el error de clasificación. Sin embargo, no siempre es posible cumplir los dos objetivos simultáneamente. Para salvar esta dificultad se plantea un problema de optimización cuya función objetivo combina ambos objetivos. Este problema de optimización resulta ser un problema de minimización cuadrático convexo [6]. En el caso que el número de objetos a clasificar es mayor que el número de atributos de cada objeto, lo que usualmente sucede, este problema tiene una única solución óptima. Al hiperplano de separación asociado a esta solución lo llamaremos el *hiperplano óptimo de separación* (HOS).

Lo descrito anteriormente corresponde al caso que existe un hiperplano de separación de las clases. En ese caso se dice que las clases son linealmente separables. Recientemente, el estudio de los SVM se ha extendido al caso de clases que no son linealmente separables mediante la introducción de las llamadas *funciones de kernel o núcleo* [17, 21] o de *variables de pérdida o slacks* [5]. Esta metodología ha sido también extendida para problemas de regresión [4].

A continuación se presenta el problema de optimización utilizado para la determinación del hiperplano de separación y se discute una solución analítica. Como resultado se obtiene la función clasificadora que discrimina los patrones de comportamiento estudiados (“fuga”–“no fuga”).

5.1.1. Obtención del Hiperplano Óptimo de Separación

Consideremos un problema de clasificación binaria para el cual ya se ha definido el conjunto de entrenamiento. En nuestro caso, los objetos a clasificar son los clientes. Supongamos que para cada cliente se han definido n variables a estudiar y que se tiene m clientes en el conjunto de entrenamiento. Entonces, cada cliente es representado por un vector característico de dimensión $n + 1$ cuyas primeras n coordenadas corresponden a las variables del estudio y la última corresponde a la clase a la que el cliente pertenece. Denotaremos estos vectores como un par (\vec{x}, y) donde $\vec{x} = (x^1, \dots, x^n)$ son las variables del estudio y la última coordenada $y \in \{-1, +1\}$ indica a qué clase pertenece el cliente. En particular, denotaremos por (\vec{x}_i, y_i) al vector característico correspondiente al cliente i (para $i = 1, \dots, m$). En este sentido, cuando quede claro del contexto, si el cliente i es representado por un vector de la forma $(\vec{x}_i, +1)$ diremos que ese cliente está en la *clase positiva* o bien, que el vector \vec{x}_i está en la clase positiva. De manera análoga, para un cliente i representado por un vector de la forma $(\vec{x}_i, -1)$ diremos que él o su vector \vec{x}_i están en la *clase negativa*.

Un supuesto inicial para este trabajo es que el conjunto de entrenamiento es linealmente separable. Es decir, existe un hiperplano en \mathfrak{R}^n que deja a todos los vectores \vec{x} asociados a los clientes de una clase de un lado del hiperplano

y al los de la otra del otro lado. Formalmente, existe un par $(\vec{\alpha}, b) \in \mathfrak{R}^{n+1}$ tal que $\vec{\alpha} \cdot \vec{x}_i + b > 0$ si el cliente i está en la clase positiva y $\vec{\alpha} \cdot \vec{x}_i + b < 0$, si está en la clase negativa.

De esta manera, dados un vector no nulo normal al hiperplano de separación $\vec{\alpha} \in \mathfrak{R}^n$ podemos definir una función de clasificación $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ por la expresión $f(\vec{x}) \equiv \vec{\alpha} \cdot \vec{x} + b$. Entonces si $f(\vec{x}_i) > 0$, el cliente i está en la clase positiva y si $f(\vec{x}_i) < 0$, i está en la clase negativa.

Para definir el margen de clasificación, consideramos las distancias d^+ y d^- . La distancia d^+ es la distancia euclídeana entre el hiperplano y la clase positiva. Es decir, la distancia entre el hiperplano y el punto en la clase positiva más cercano a este. De manera análoga, pero respecto de la clase negativa se define d^- . Una vez determinadas estas distancias, definimos el *margen de separación* (del hiperplano) como la suma $d^+ + d^-$.

5.2. Construcción del Problema de Optimización

De la discusión anterior sabemos que para clasificar correctamente a todos los clientes en el conjunto de entrenamiento debemos imponer las restricciones:

$$\begin{aligned} \vec{\alpha} \cdot \vec{x}_i + b &> 0 \text{ para los clientes } i \text{ en la clase positiva} \\ \text{y} & \\ \vec{\alpha} \cdot \vec{x}_i + b &< 0 \text{ para los clientes } i \text{ en la clase negativa.} \end{aligned} \tag{2}$$

No es una buena práctica incluir desigualdades estrictas como restricciones en problemas de optimización. Por lo tanto, sería importante poder encontrar condiciones equivalentes a las anteriores pero que sean del tipo “menor o igual” o “mayor o igual”.

Recordando que si un hiperplano está definido por el par $(\vec{\alpha}, b)$, entonces cualquier par de la forma $(\lambda\vec{\alpha}, \lambda b)$ con $\lambda > 0$ también define el mismo hiperplano, podemos que si $\vec{\alpha} \cdot \vec{x} + b > 0$ para un par particular $(\vec{\alpha}, b)$, entonces existe un par $(\vec{\alpha}_i, b_i)$ que define el mismo hiperplano y tal que $\vec{\alpha}_i \cdot \vec{x} + b_i \geq 1$. Para la clase negativa, se puede razonar de manera análoga. Como el conjunto de entrenamiento es finito, si existe un par $(\vec{\alpha}, b)$ que satisface las condiciones (2) para todos los clientes del conjunto de entrenamiento, entonces existe un par $(\vec{\alpha}', b')$ que define el mismo hiperplano tal que el lado izquierdo de las expresiones en (2) tiene valor absoluto mayor o igual a 1. De esta manera, las condiciones de la ecuación (2) pueden ser incluidas en el problema de optimización en la forma:

$$\begin{aligned} \vec{\alpha} \cdot \vec{x}_i + b &\geq 1 \text{ para los clientes } i \text{ en la clase positiva} \\ \text{y} & \\ \vec{\alpha} \cdot \vec{x}_i + b &\leq -1 \text{ para los clientes } i \text{ en la clase negativa.} \end{aligned}$$

Los puntos (clientes) para los cuales una de estas restricciones es activa, es decir $|\vec{\alpha} \cdot \vec{x}_i + b| = 1$ tienen especial importancia. En particular, estos puntos

yacen exactamente en los llamados *hiperplanos canónicos* que están definidos por las ecuaciones [7]:

$$\begin{aligned} \vec{\alpha} \cdot \vec{x} + b &= +1 \\ \vec{\alpha} \cdot \vec{x} + b &= -1. \end{aligned} \tag{3}$$

Denotaremos por x^+ a un cliente cualquiera que esté en el hiperplano canónico de la clase positiva (lado derecho igual a +1) y por x^- a un cliente cualquiera que esté en el hiperplano canónico de la clase negativa. El hiperplano definido por el par $(\vec{\alpha}, b)$ es paralelo a los hiperplanos canónicos.

Con esta notación, el margen de separación del hiperplano definido por el par $(\vec{\alpha}, b)$, y denotado por γ , puede ser expresado como [7]:

$$\gamma = \frac{1}{2} \left[\left(\frac{\vec{\alpha} \cdot x^+}{\|\vec{\alpha}\|} \right) - \left(\frac{\vec{\alpha} \cdot x^-}{\|\vec{\alpha}\|} \right) \right] \tag{4}$$

Manipulando algebraicamente la expresión anterior se obtiene:

$$\gamma = \frac{1}{2\|\vec{\alpha}\|} [(\vec{\alpha} \cdot x^+) - (\vec{\alpha} \cdot x^-)] = \frac{1}{\|\vec{\alpha}\|} \tag{5}$$

Observemos que la distancia entre el hiperplano definido por el par $(\vec{\alpha}, b)$ a cada uno de los hiperplanos canónicos es igual a $1/\|\vec{\alpha}\|$. Además, los puntos más cercanos a este hiperplano en el conjunto de entrenamiento están en los hiperplanos canónicos. Por lo tanto, el margen de separación de este hiperplano es igual a $2/\|\vec{\alpha}\|$. En la figura 9 se ilustra esta situación: H_1 y H_2 son los hiperplanos canónicos.

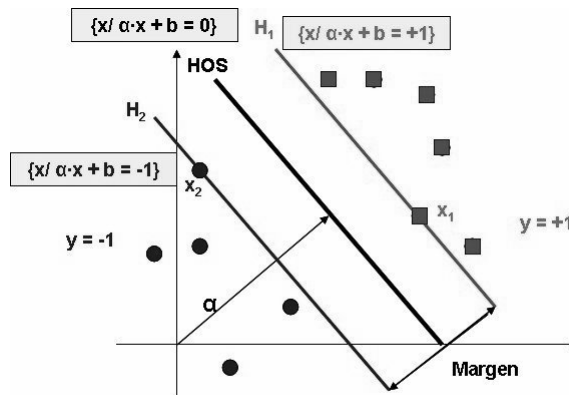


Figura 9: Hiperplanos canónicos de separación.

Se desea maximizar este margen y esto es equivalente a minimizar la norma euclidiana de $\vec{\alpha}$. De esta manera el problema de optimización que vamos a resolver es el siguiente [7]:

$$\begin{aligned}
 & \underset{\vec{\alpha}, b}{\text{máx}} && \frac{1}{2} \|\vec{\alpha}\|^2 \\
 & \text{s.a.} && \\
 & && y_i \cdot (\vec{x}_i \cdot \vec{\alpha} + b) - 1 \geq 0 \quad \text{para } i = 1, \dots, m.
 \end{aligned} \tag{P}$$

Resolviendo este problema se obtiene un vector $\vec{\alpha}^*$ óptimo que permite calcular el margen geométrico máximo de un hiperplano de separación, $\gamma^* = 2/\|\vec{\alpha}^*\|$.

En la próxima sección, se resuelve analíticamente este problema de optimización y se describe un sistema de ecuaciones que permite calcular el vector óptimo $\vec{\alpha}^*$ y el lado derecho b .

5.2.1. Resolución Analítica del Problema de Optimización

El problema de optimización planteado en la sección anterior es un problema cuadrático convexo. En el caso que nos interesa, este problema tiene una única solución óptima global. Esta solución puede ser obtenida encontrando una solución para las condiciones de optimalidad de primer orden, conocidas como las condiciones de Karush, Kuhn y Tucker [7]. Para esto definimos el *Lagrangiano* de (P),

$$L_P \equiv \frac{1}{2} \|\vec{\alpha}\|^2 - \sum_{i=1}^m \beta_i y_i (\vec{x}_i \cdot \vec{\alpha} + b) + \sum_{i=1}^m \beta_i \tag{6}$$

donde los β_i son los multiplicadores de Lagrange asociados a las restricciones de (P). A partir de este Lagrangiano podemos plantear las condiciones de optimalidad [2]:

$$\frac{\partial L_P}{\partial \alpha_j} = \alpha_j - \sum_{i=1}^m \beta_i y_i x_i^j = 0 \quad j = 1, \dots, n \tag{7}$$

$$\frac{\partial L_P}{\partial b} = - \sum_{i=1}^m \beta_i y_i = 0 \tag{8}$$

$$y_i (\vec{x}_i \cdot \vec{\alpha} + b) - 1 \geq 0 \quad i = 1, \dots, m \tag{9}$$

$$\beta_i \geq 0 \quad i = 1, \dots, m \tag{10}$$

$$\beta_i (y_i (\vec{\alpha} \cdot \vec{x}_i + b) - 1) = 0 \quad i = 1, \dots, m \tag{11}$$

De las condiciones (7) y (8) se obtienen, en el óptimo, las siguientes relaciones:

$$\vec{\alpha} = \sum_{i=1}^m \beta_i y_i \vec{x}_i \tag{12}$$

$$\sum_{i=1}^m \beta_i y_i = 0 \tag{13}$$

Al reemplazar estas expresiones en la ecuación (6) se puede llevar el Lagrangiano a la forma:

$$L_D \equiv \sum_{i=1}^m \beta_i - \frac{1}{2} \sum_{i,s=1}^m \beta_i \beta_s y_i y_s \vec{x}_i \cdot \vec{x}_s \quad (14)$$

A partir de esta expresión, aplicando dualidad lagrangiana se puede obtener un problema dual conocido como el Dual de Wolfe [2, 8] de (P):

$$\begin{aligned} \max_{\beta} \quad & L_D \equiv \sum_{i=1}^m \beta_i - \frac{1}{2} \sum_{i,s=1}^m \beta_i \beta_s y_i y_s \vec{x}_i \cdot \vec{x}_s \\ \text{s.a.} \quad & \sum_{i=1}^m \beta_i y_i = 0 \\ & \beta_i \geq 0 \quad \text{para } i = 1, \dots, m. \end{aligned} \quad (D)$$

En el caso de problemas de optimización convexos regulares, como el problema (P), el Dual de Wolfe (D) es un dual fuerte. La resolución se reduce, entonces, a obtener los valores óptimos para los multiplicadores β_i . Una vez conocidos estos, los valores óptimos de las variables primales $\vec{\alpha}$ se obtienen de la ecuación (12).

A pesar de que el término b o *bias* no aparece explícitamente en la formulación, puede ser obtenido a partir de las condiciones de optimalidad [7]:

$$b^* = -\frac{1}{2} \left(\max_{y_i=-1} (\vec{\alpha}^* \cdot \vec{x}_i) + \max_{y_i=+1} (\vec{\alpha}^* \cdot \vec{x}_i) \right) \quad (15)$$

Resumiendo, para encontrar el par $(\vec{\alpha}^*, b^*)$ que define el hiperplano de separación óptimo se resuelve el problema dual (D) obteniéndose los multiplicadores $\vec{\beta}^*$ óptimos. Utilizando estos valores en las expresiones (12) y (15) se obtienen $\vec{\alpha}^*$ y b^* , respectivamente.

6. Resultados del Modelo Predictivo

El modelo desarrollado con la técnica SVM fue comparado con el sistema de predicción de fugas que poseía la institución. Este sistema utilizaba como técnica de modelación una red neuronal multicapa (MLP). Las comparaciones entre el modelo antiguo y el modelo propuesto, se realizaron en base a una muestra común, utilizando el mismo número de atributos (variables) y los mismos conjuntos de entrenamiento y test.

El cuadro 2 muestra un resumen de los resultados obtenidos por el modelo SVM y la red MLP, en términos de porcentaje de acierto.

Porcentajes de Acierto				
Modelo	Entrenamiento (%)	Varianza	Test(%)	Varianza
SVM	89.41	0.01	87.91	0.01
RED MLP	80.40	0.10	77.47	0.20

Cuadro 2: Comparación entre el Modelo SVM y Red Neuronal MLP.

Tanto para el conjunto de entrenamiento, como para el conjunto de test, el modelo SVM fue superior a la red MLP.

Otro punto a destacar es que los resultados arrojados por la red MLP poseen una mayor varianza, lo que va en desmedro de la estabilidad de los resultados.

También se comparó la efectividad de ambas técnicas, en dos meses consecutivos, respecto a la cartera total de clientes. Se aplicaron ambos modelos a un conjunto formado por 60.000 clientes, en donde se seleccionaron para cada modelo los 5.000 clientes que poseían mayores tendencias a la fuga.

Cabe destacar que los conjuntos de clientes con tendencias a la fuga generados por ambas técnicas pueden ser diferentes, ya que puede pasar que para un modelo un cliente tenga claros síntomas de fuga y para el otro no.

El cuadro 5.2.1 se detallan los aciertos para cada conjunto de clientes: por ejemplo, en el mes 1 se observa que al comparar los 20 clientes con mayores síntomas de fuga por cada modelo, SVM identifica 10 casos, mientras que la red MLP no identifica ninguno de los casos de fuga.

Se observa la superioridad en la predicción de fuga de los clientes en ambos meses por parte del modelo propuesto (SVM), siendo esta superioridad acentuada en los primeros segmentos de clientes seleccionados (los de más alta tendencia).

Clientes con Mayores Tendencias a la Fuga							
Mes 1							
Modelo	20	50	100	500	1.000	2.000	5.000
SVM	10	15	24	53	80	122	274
Red MLP	0	3	3	10	17	51	251
Mes 2							
SVM	6	10	20	40	68	118	277
Red MLP	1	3	5	20	36	83	285

Cuadro 3: Comparación de aciertos.

7. Conclusiones y Trabajo Futuro

7.1. Conclusiones del Estudio

El modelo de predicción de fugas es una herramienta importante de apoyo a la hora de decidir cuáles de los clientes de la cartera poseen una mayor tendencia a la fuga. Al identificar de mejor forma a estos clientes es posible: (1) aumentar las utilidades y la rentabilidad del negocio, ya que da la posibilidad de retener clientes fugitivos y mantenerlos como clientes captando sus flujos futuros; (2) focalizar los recursos productivos sobre los segmentos que realmente necesitan de acciones de retención fuertes aumentando la eficiencia y efectividad de las políticas de retención; (3) generar un mejoramiento y fortalecimiento de la relación entre el cliente y la institución, al anticiparse a sus necesidades, así como hacer menos sensible al cliente frente a las campañas de marketing de la competencia; (4) disminuir el riesgo total de la cartera, al reducir la necesidad de atraer nuevos clientes potencialmente más riesgosos.

Respecto a la técnica utilizada para la modelación, se reafirma su fortaleza en términos de predicción alcanzando diferencias cercanas a los 10 puntos porcentuales respecto a los resultados obtenidos por el sistema que usaba la institución (Red MLP). En términos de estabilidad, el modelo propuesto entregó resultados con una menor varianza, generando resultados más robustos.

7.2. Trabajo Futuro

Las nuevas líneas de investigación y trabajos futuros se centran en los siguientes aspectos:

1. Desarrollar nuevos estudios sobre la identificación de atributos relevantes para la clasificación. Un enfoque en esta línea es la penalización de los atributos en la función objetivo.
2. Desarrollar nuevas formulaciones que minimicen el número de puntos sobre los hiperplanos canónicos. Algunas cotas teóricas dependen de este número, siendo más ajustadas mientras menos puntos haya en los hiperplanos canónicos.

Agradecimientos Los autores les agradecen al Núcleo Científico Milenio “Sistemas Complejos de Ingeniería” P04-066-F y al proyecto Fondecyt N°1040926, por el apoyo brindado en la elaboración y financiamiento de este estudio.

Referencias

- [1] A. Athanassopoulos. Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of Business Research*, 47(3):197–207, 2000.
- [2] M. Bazaraa, H. Sherali, and C. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley, New York, second edition, 1993.
- [3] C.B. Bhattacharya. When customers are members: Customer retention in paid membership context. *Journal of the Academy of Marketing Science*, 26(1):31–44, 1998.
- [4] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
- [5] C. Cortés and V. Vapnik. Support vector networks. *Journal of Machine Learning*, 20: 273–297, 1995.
- [6] N. Cristianini and R. Holloway. *Support Vector and Kernel Methods*. Springer-Verlag, Berlin, Heidelberg, 2003.
- [7] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge, UK, Cambridge University, 2000.
- [8] R. Fletcher. *Practical Methods of Optimization*. John Wiley and Sons, New York, 1987.
- [9] J. Ganesh, M.J. Arnold, and K.E. Reynolds. Understanding the customer base of service providers: An examination of the differences between switchers and stayers. *Journal of Marketing*, 64(3):65–87, 2000.
- [10] M.A. Jones, D.L. Mothersbaugh, and S.E. Beatty. Switching barriers and repurchase intentions in services. *Journal of Retailing*, 72(2):259–374, 2000.
- [11] M.A. Jones, D.L. Mothersbaugh, and S.E. Beatty. Why customers stay: measuring the underlying dimensions of services switching costs and managing their differential strategic outcomes. *Journal of Business Research*, 55(6):441–450, 2002.
- [12] B. Mittal and W.A. Kamakura. Satisfaction, repurchase intent, and repurchase behavior: investigating the moderating effect of customer characteristics. *Journal of Marketing Research*, 1(131–142), 2000.

- [13] M. Paulin, J. Perrien, R.J. Ferguson, and A.M.A. Salazar. Relational norms and client retention: external effectiveness of commercial banking in canada and mexico. *International Journal of Bank Marketing*, 16(1):24–31, 1998.
- [14] H.C. Pohani, K.R. Lane, and S. Pofcher. Format renewal in banks – it’s not easy. *McKinsey Quarterly*, 2:148–154, 1998.
- [15] E. Ramusson. Complaints can build relationships. *Sales and Marketing Management*, 151(9):89–90, 1999.
- [16] F. Reichheld and E. Sasser. Zero defections: Quality comes to services. *Harvard Business Review*, 1990:105–111, September–October, 2000.
- [17] B.Schölkopf, K. Sung, C. Burgues, J.C. Girosi, P. Niyogui, and V. Vapnik. Comparing support vector machine with gaussian kernels to radial basis function classifiers. *IEEE Transaction on Signal Processing*, 45(11):2758–2765, 1997.
- [18] D. Van den Poel and Bart Larivière. Customer attrition analysis for financial services using proportional hazard models. Technical Report B–9000, Department of Marketing, Ghent University, Hoveniersberg 24, 2003.
- [19] V. Vapnik. *Estimation of dependences based on empirical data*. Springer-Verlag, New York, 1982.
- [20] V. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, New York, 1995.
- [21] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, New York, 1998.