ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

RENTABILIZACIÓN DE CLIENTES POR MEDIO DEL APRENDIZAJE AUTOMÁTICO

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERA MATEMÁTICA

PROYECTO DE INVESTIGACIÓN

RUTH MARICELA TIPANTUÑA JARRÍN
ruth22maricela@gmail.com

DIRECTORA: SANDRA E. GUTIÉRREZ POMBOSA, PH.D
sandra.gutierrez@epn.edu.ec

QUITO, NOVIEMBRE 2018

DECLARACIÓN

Yo RUTH MARICELA TIPANTUÑA JARRÍN, declaro bajo juramento que el trabajo aquí escrito es de mi autoría, que no ha sido previamente presentado para ningún grado o calificación profesional y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la Normatividad Institucional vigente.

Ruth Maricela Tipantuña Jarrín

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por RUTH MARICELA TIPANTUÑA JARRÍN bajo mi supervisión.

Sandra Gutiérrez, Ph.D
Director del Proyecto

AGRADECIMIENTOS

Agradezco a mis padres, Luz y Santiago. A mis hermanos Marco, Mónica, Hugo y Adrián.
Además deseo agradecer a todos mis amigos y a quienes me ayudaron a realizar este trabajo, muchas gracias!

DEDICATORIA

Este trabajo lo dedico a mis padres y hermanos

Índice general

1. Introducción **1**
1.1. Estado del arte de la rentabilidad de clientes 1
1.2. Deserción de clientes 3
1.3. Deserción de clientes en banca 6

2. Aprendizaje automático **11**
2.1. Preliminares 11
2.2. Minería de datos 11
2.2.1. Minería de datos en el marco de CRM 12
2.2.2. Proceso de descubrimiento del conocimiento 13
2.2.3. Modelos Predictivos / Supervisados y técnicas 16
2.2.4. Evaluación de Modelos de Clasificación 19
2.2.5. Modelos no supervisados y técnicas 20
2.3. Machine Learning 21
2.3.1. Pasos a seguir en el aprendizaje 24

3. MÁQUINAS DE SOPORTE VECTORIAL (SUPPORT VECTOR MACHINE, SVM) **25**
3.1. Minimización del Riesgo Empírico 25
3.2. Minimización de Riesgo Estructural 27
3.3. Propiedades de las máquinas de soporte vectorial 28
3.4. Clasificador de margen máximo. 29
3.4.1. Hiperplano de separación 29
3.4.2. Clasificación usando un hiperplano 30
3.4.3. Hiperplano de separación óptimo 31
3.5. Máquina de soporte vectorial 33
3.5.1. Caso separable, SVM de margen rígido (SVM lineal) 33
3.5.2. Caso cuasi separables, SVM de margen suave o clasificador de vector de soporte 36
3.5.3. Caso no separable linealmente 38
3.6. El problema de desbalance en los datos 40
3.6.1. Enfoques para afrontar el desbalance de datos 42
3.6.2. Métricas de evaluación más apropiadas para el caso desbalanceado 44
3.6.3. Definición de métricas para clasificación en el caso desbalanceado 46

4. MODELO DE DESERCIÓN **48**
4.1. Datos y supuestos 48
4.1.1. Comprensión del negocio 48
4.1.2. Comprensión de los datos 50
4.1.3. Definición de deserción 50
4.1.4. Propuesta de un marco de predicción de abandono dinámico . 51
4.1.5. Selección de variables 55
4.2. Representación 56
4.3. Método y estimación 57
4.3.1. Tiempo de predicción 57
4.3.2. Modelamiento con regresión logística 57
4.3.3. Modelamiento con SVM 58
4.4. Evaluación 60
4.5. Selección del modelo 60
4.6. Aplicación al cálculo del CLV 65
4.7. Gestión de la rentabilidad tras una campaña de retención 69

5. Conclusiones y Recomendaciones **71**
5.1. Conclusiones 71
5.2. Recomendaciones 72

Anexos **73**

A. Tablas de variables significativas mediante árboles de decisión **74**

B. Código regresión logística en R **78**

C. Código SVM en R **80**

D. Coeficientes del hiperplano de separación **83**

E. Resultados obtenidos con regresión logística **94**

F. Divergencia Kullback-Leibler **96**

G. Comportamiento de un cliente desertor **97**

Índice de cuadros

1.1. Tipos de CRM. Fuente: [Buttle, 2009] 2

3.1. Matriz de confusión para la evaluación de clasificadores 45

4.1. Muestras obtenidas tras realizar submuestreo 51
4.2. Variables consideradas en el modelo logístico. Elaborado por: Autor . 58
4.3. Matrices de confusión obtenidas con regresión logística en tres conjuntos de prueba 59
4.4. Matrices de confusión obtenidas con regresión logística en tres conjuntos de entrenamiento 59
4.5. Matrices de confusión obtenidas con SVM kernel lineal en tres conjuntos de prueba 61
4.6. Matrices de confusión obtenidas con SVM kernel lineal en tres conjuntos de entrenamiento 61
4.7. Matrices de confusión obtenidas con SVM kernel radial en tres conjuntos de prueba 62
4.8. Matrices de confusión obtenidas con SVM kernel radial en tres conjuntos de entrenamiento 62
4.9. Matrices de confusión obtenidas con SVM kernel polinomial en tres conjuntos de prueba 63
4.10. Matrices de confusión SVM kernel polinomial en tres conjuntos de entrenamiento 63
4.11. Resultados del modelo en el conjunto de prueba 64
4.12. Resultados del modelo en el conjunto de entrenamiento 64
4.13. Valores AUC obtenidos en los conjuntos de entrenamiento 68

A.1. Variables significativas cuadro 1. Elaborado por: Autor 74
A.2. Variables significativas cuadro 2 75
A.3. Variables significativas cuadro 3 76
A.4. Variables significativas cuadro 4 77

D.1. Pesos del hiperplano de separación, tabla 1. Elaborado por: Autor. . 84
D.2. Pesos del hiperplano de separación, tabla 2. 85
D.3. Pesos del hiperplano de separación, tabla 3. 86
D.4. Pesos del hiperplano de separación, tabla 4. 87
D.5. Pesos del hiperplano de separación, tabla 5. 88
D.6. Pesos del hiperplano de separación, tabla 6. 89
D.7. Pesos del hiperplano de separación, tabla 7. 90
D.8. Pesos del hiperplano de separación, tabla 8. 91

D.9. Pesos del hiperplano de separación, tabla 9. 92
D.10.Add caption 93

E.1. Resultados con regresión logística. Elaborado por: Autor 95

Índice de figuras

3.1. Observaciones que pertenecen a dos clases, que se muestran en azul y en morado. Además tres hiperplanos de separación de entre muchos posibles 30
3.2. Hiperplano de separación 31
3.3. En este gráfico hay dos clases de observaciones, que se muestran en azul y en morado. El hiperplano de margen máximo se muestra como una línea continua. El margen es la distancia desde la línea continua a cualquiera de las líneas punteadas. Los dos puntos azules y el punto violeta que se encuentran en las líneas punteadas son los vectores de soporte, y la distancia desde esos puntos al hiperplano se indica con flechas. 32
3.4. En este caso, las dos clases no son separables por un hiperplano, por tanto no se puede usar el clasificador de margen máximo. 37

4.1. Línea de tiempo fija. Fuente [Bhambri, 2013] 54
4.2. Línea de tiempo dinámica. Fuente [Bhambri, 2013] 54
4.3. Curva ROC modelo SVM kernel: radial. Muestra 1. Elaborado por: Autor. 65
4.4. Curva ROC modelo SVM kernel: radial. Muestra 2. Elaborado por: Autor. 66
4.5. Curva ROC modelo SVM kernel: radial. Muestra 3. Elaborado por: Autor. 67

G.1. Total corrientes saldo contable para un cliente desertor. Elaborado por: Autor 97
G.2. Total corrientes saldo promedio para un cliente desertor. Elaborado por: Autor 98

RESUMEN

Este trabajo se centra en predecir la deserción de un cliente en un tiempo determinado en el sector bancario con el fin establecer una estrategia de retención. Para lograr este objetivo es importante entender el contexto de negocio en que se desarrolla el problema y el tipo de deserción que se predecirá. Puesto que se desea establecer una estrategia de retención, la predicción se realiza con 3 meses de anticipación al evento de deserción. Las herramientas para lograr esto son: regresión logística y máquinas de soporte vectorial con el fin de verificar la exactitud del aprendizaje automático frente a una técnica tradicional. Puesto que el problema de deserción posee desbalance en los datos, es decir, la diferencia de la proporción entre desertores y no desertores es muy grande se usó submuestreo para afrontar este problema. Una vez calculada la probabilidad de deserción se calculará una aproximación del valor del cliente para determinar si vale la pena o no retenerlo.

Palabras clave: SVM (Máquinas de soporte vectorial), Deserción, CRM (Gestión de relación con el cliente), CLV (Valor del tiempo de vida del cliente), submuestreo, regresión logística.

ABSTRACT

This work focuses on predicting the desertion of a client at a certain time in the banking sector in order to establish a retention strategy. To achieve this goal it is important to understand the business context in which the problem develops and the type of defection that will be predicted. Since it is desired to establish a retention strategy, the prediction will be made 3 months before the attrition event. The tools to achieve this are: logistic regression and vector support machines in order to verify the accuracy of machine learning versus a traditional technique. Since the problem of desertion has an imbalance in the data, that is, the difference in the proportion between dropouts and non-dropouts is very large, undersampling was used to deal with this problem. Once the probability of desertion is calculated, an approximation of the value of the client will be calculated to determine if it is worthwhile or not to retain it.

Keywords: SVM (Support Vector Machine), Churn (Change and turn), CRM (Customer Relationship Management), CLV (Customer Lifetime Value), undersampling, logistic regression.

Capítulo 1

Introducción

La rentabilización de clientes por medio del aprendizaje automático involucra conceptos claves como: gestión de relación con los clientes (CRM: Customer Relationship Management), máquinas de soporte vectorial (SVM: Support Vector Machine) y el valor del tiempo de vida del cliente (CLV: Customer Lifetime Value).

Se parte de la estrategia de la gestión de las relaciones entre una empresa y los clientes pues indica el proceso que podría seguirse con el fin de conseguir que un cliente llegue a ser rentable. Un paso clave en esta estrategia es la retención del cliente, para esto es necesario predecir si un cliente desertará o no en un tiempo determinado, este paso se puede abordar como un problema de clasificación binario usando SVM. Una vez que se predice la deserción es importante saber si vale o no la pena emprender una campaña de retención hacia un cliente, esto se puede saber calculando el valor del tiempo de vida del cliente CLV.

1.1. Estado del arte de la rentabilidad de clientes

En vista que este trabajo se centra principalmente en rentabilizar clientes del sector bancario se parte de una descripción de conceptos relacionados con estrategias de negocios que permiten lograr este objetivo, además que ayudará a tener una idea sobre cómo actuar una vez que se han obtenido los resultados en el capítulo 4.

Hoy en día los consumidores están mejor informados, son más exigentes en los productos y servicios que necesitan, además están más familiarizados con la tecnología. Por el lado de la oferta del mercado, existe un incremento significativo de la competencia, con muy poca diferenciación de productos y servicios, permitiendo que los consumidores tengan pocas opciones claras. El resultado final de la combinación de estos factores es que los consumidores han elevado su poder adquisitivo a niveles más altos.

Hay cuatro cuestiones principales que las organizaciones deben abordar para satisfacer a los consumidores exigentes:

- Retener a los clientes e impedir que se dirijan a la competencia.

- Determinar qué productos y servicios agrupar para aumentar la rentabilidad del cliente.
- Tratar a los clientes de manera individual.
- Implementar soluciones tecnológicas que logren objetivo de la empresa, [Sharp, 2002].

El implementar una filosofía de negocio que se enfoque en satisfacer a los clientes traerá beneficios como el aumento de los ingresos, reducción de costos lo que resultará en una base de clientes leales.

El área que se encarga de manejar cuestiones relacionados con los clientes es la *gestión de la relación con el cliente* (CRM: Customer Relationship Management), este término ha estado en uso desde principios de los años noventa. Los diferentes tipos de CRM se resumen en el Cuadro 1.1.

Cuadro 1.1: Tipos de CRM. Fuente: [Buttle, 2009]

Tipo de CRM	**Descripción**
Estratégico	Este tipo de CRM es una estrategia centrada en el cliente que tiene como objetivo ganar y mantener a los clientes rentables.
Operacional	El CRM Operacional se centra en la automatización de los procesos orientados al cliente como la venta, la comercialización y el servicio al cliente.
Analítico	Se centra en la minería inteligente de datos relacionados con los clientes con fines estratégicos o tácticos.
Colabrativo	El CRM colaborativo aplica la tecnología a través de fronteras organizacionales con el fin de optimizar el valor de la empresa, socios y clientes. Permite a organizaciones separadas alinear sus esfuerzos para atender a los clientes con mayor eficacia compartiendo información valiosa a lo largo de la cadena de suministro.

Al identificar los atributos principales de la clasificación CRM e integrarlos se obtiene la siguiente definición: *"CRM es la principal estrategia de negocio que integra procesos y funciones internas, así como redes externas, para crear y entregar valor a clientes objetivo en un determinado perfil. Se basa en la alta calidad de datos relacionados con el cliente y habilitado por la tecnología de la información"*, [Buttle, 2009].

Esta estrategia se ha implementado en varios sectores comerciales tales como: bancos, fabricantes de automóviles, empresas de alta tecnología y fabricantes de bienes de consumo.

- Los bancos necesitan aplicar CRM pues les permite gestionar la tasa de deserción de clientes. Las técnicas de minería de datos se pueden utilizar

para identificar qué clientes son propensos a desertar, qué se puede hacer para recuperarlos, qué clientes son propensos a ciertas ofertas y la mejor forma de comunicarlos.

- Los fabricantes de automóviles venden a través de las redes del distribuidor o concesionario. Puesto que tienen poco contacto con el usuario, propietario o conductor, usan CRM para desarrollar relaciones rentables con sus redes de distribución.
- Las empresas de alta tecnología que fabrican productos complejos que generalmente son vendidas por las organizaciones asociadas. Por ejemplo, pequeños desarrolladores de software innovadores tradicionalmente se asocian con empresas como IBM para la distribución y ventas. En este caso también se han desarrollado relaciones automatizadas con los proveedores, de modo que puedan cumplir con los niveles de inventario.

Sin duda, las empresas generan mejores resultados cuando manejan su base de clientes con el fin de identificar, adquirir, retener y satisfacer las necesidades de clientes rentables, los cuales son los objetivos clave de muchas estrategias de CRM, [Buttle, 2009].

Sin embargo, la campaña de marketing directo más sofisticada, visionaria y estratégica fracasará en el primer obstáculo si se aplica a datos de clientes que no son relevantes. La extracción y limpieza de datos es un ejercicio costoso que consume mucho tiempo, por lo que es vital para hacerlo bien la primera vez [Sharp, 2002].

Este trabajo de titulación se centra principalmente en la etapa de retener y satisfacer las necesidades de clientes rentables.

Una solución de CRM bien definida y basada en un sistema de almacenamiento de datos, permite a las empresas capturar y analizar las interacciones y transacciones de los clientes para reducir el abandono de los mismos así como ayudar a las empresas a comprender mejor sus necesidades y cambios en los patrones de compra, además permite construir relaciones a largo plazo de valor para ellos [Sharp, 2002].

A continuación se describe por qué centrarse en la deserción de un cliente es importante para aumentar el valor de un cliente y por tanto generar beneficios para una empresa.

1.2. Deserción de clientes

La deserción de clientes se ha convertido en un problema importante y es uno de los principales desafíos que muchas empresas en todo el mundo están afrontando [Chandar and Krishna, 2006].

Hoy en día, el análisis de deserción se aplica a diferentes sectores tales como mercados financieros, proveedores de servicios de correo electrónico, minoristas de productos industriales, telecomunicaciones. El objetivo de esta técnica consiste en detectar a los clientes con mayor riesgo de deserción y en la implementación de acciones de marketing para incrementar la lealtad, reduciendo la probabilidad de que se produzca una deserción.
El análisis de deserción es una técnica que ayuda a comprender el comportamiento del cliente y predecir el momento en que ocurrirá el evento deserción, el típico evento de abandono está vinculado al hecho de que un cliente deje de usar un servicio específico para pasar a la competencia [Birtolo and Ritrovato, 2013].

Según [Seufert, 2013], adquirir un nuevo usuario implica incertidumbre. Una persona sólo está dentro del alcance del análisis de un producto una vez que la persona ha interactuando con el producto por primera vez. Un usuario fuera de ese ámbito, en quien no se tiene una comprensión acerca del uso del producto se ubica en un espectro muy amplio de perfiles de comportamiento, la mayoría de los cuales no conducen a contribuciones directas de ingresos hacia producto. Por lo tanto, un usuario con un valor del tiempo de vida de $ 0 que está comprometido con el producto es aún más valioso que un usuario recién adquirido, dadas las siguientes afirmaciones:

- Los datos del producto y el usuario existente tienen valor en términos de su utilidad para la creación de modelos predictivos sobre los usuarios futuros. Una vez más, el nuevo usuario es un misterio mientras que los datos sobre un usuario que ha tenido una larga relación con la empresa es más valioso que los datos de un usuario de corta duración.

- El nuevo usuario debe ser adquirido, lo que representa un gasto. Incluso si el usuario ya existente fue adquirido a través de una campaña pagada en el pasado, ese costo no es relevante para su estado actual. Por lo general, en software, el costo de mantener un único usuario versus el costo de adquirir un nuevo usuario es insignificante.

Por lo tanto, la estrategia de la industria inalámbrica y otras han pasado de una orientación de adquisición a una política de retención. Predecir la deserción hace que esta estrategia sea factible [Lemmens and Croux, 2006].
Las ventajas relacionadas con la lealtad de los clientes justifican la creciente preocupación de las empresas por el abandono de sus clientes” [Migueis and e Cunha, 2012].

Los programas de lealtad pueden crear diferentes tipos de barreras al abandono, las cuáles pueden ser económicas, en el caso que los clientes pierden ventajas, por ejemplo puntos si ellos cambian el producto o de proveedores, las barreras también pueden ser psicológicas, sociológicas que realzan el compromiso de los clientes y la confianza en la organización de acuerdo a [Morgan and Hunt, 1994], lo cual reforzará los efectos de un programa de lealtad más allá de aspectos económicos.
Al implementarse estos programas los consumidores pueden apreciar que las recompensas los hacen sentir como clientes preferidos y por lo tanto se identificarán

más con la compañía [Oliver, 1999], de manera que, con una relación interactiva, de alta calidad, a largo plazo conduce a una mayor confianza, compromiso y lealtad [Meyer-Waarden, 2007].

Para medir la relación de un cliente con una empresa, dos conceptos clave pueden utilizarse como parte de un enfoque analítico, el **valor de la vida del cliente** y la **deserción** [Lima and Baesens, 2009].

En este trabajo se utilizará la medida del *valor de la vida del cliente* para medir la rentabilidad de un cliente con una probabilidad alta de deserción y de esa manera saber si vale la pena retenerlo emprendiendo una campaña de retención.

El valor del tiempo de vida del cliente (Customer Lifetime Value, CLV) representa el valor actual de los beneficios esperados menos los costos de inicialización, mantenimiento y desarrollo de la relación con el cliente, esto implica elementos tales como costos, ingresos, una tasa de descuento y un horizonte de tiempo. Otro aspecto importante es la tasa de retención o la probabilidad de deserción la cuál es necesaria al considerar la evaluación del valor del tiempo de vida del cliente, en lugar de sólo mirar la historia de ganancias pasadas [Lima and Baesens, 2009].

Ampliando más el concepto de valor del tiempo de vida del cliente, se puede decir que es un indicador de la futura rentabilidad esperada de los clientes para una organización. Esta cuestión es de interés particular para las organizaciones que se proponen maximizar el valor de una relación con el cliente sobre un período de tiempo particular. En contraste con la práctica habitual de establecer objetivos de rendimiento separadas para las diferentes funciones de la empresa, que pueden entrar en conflicto entre sí, el CLV ofrece una herramienta de ayuda a la decisión integral que da “el mismo enfoque a lo largo de las diferentes áreas de toma de decisiones de la organización” [Audzeyeva and Schenk-Hopp, 2012].

El analizar la retención del cliente, lealtad, y la deserción en el contexto de valor de vida de cliente (CLV) es primordial pues una empresa tendrá el sentido de qué realmente está perdiendo debido a la deserción y la escala de los esfuerzos que serían apropiados para una campaña de retención. El análisis del valor del cliente con el pronóstico de deserción ayudarán al objetivo de programas de marketing que tienen que ver con centrarse grupos más específicos de clientes [Lima and Baesens, 2009].

El análisis del CLV también puede ayudar a la empresa en su personalización de productos y servicios. Esta comprensión del valor del cliente ayuda a la empresa a centrarse en los clientes con ingresos productivos y en el rendimiento del segmento de clientes con potenciales impactos negativos a los ingresos. Y por último, el valor de la vida del cliente no es un valor fijo, puede ser influenciado por los esfuerzos de marketing y lograr que este valor aumente [Mutanen and Nousiainen, 2006].

El considerar el valor del tiempo de vida del cliente (CLV) y la deserción es importante, pues mientras que las actividades de marketing de base de datos tal como cross-selling y up-selling[1], recompensa de frecuencia o programas a nivel de cliente están centrados en desarrollar las relaciones con los mismos, hay siempre el temor que en medio de estos esfuerzos el cliente decida dejar la empresa, es decir, desertar. Por tanto es necesario controlar la deserción del cliente, centrándose en la gestión proactiva de la misma, donde la empresa se pone en contacto con el cliente de manera previsiva y proporciona un servicio o un incentivo diseñado para evitar que el cliente deserte [Blattberg and Neslin, 2008].

1.3. Deserción de clientes en banca

De acuerdo a uno de los objetivos del plan de trabajo, en esta sección se describe la deserción de clientes en el contexto bancario.

La mayoría de las instituciones financieras saben implícitamente que algunos clientes son más rentables que otros, pero lamentablemente muchos siguen tratando a todos los clientes de la misma manera. Algunos bancos han pensado por ejemplo en la regla de 80/20 esto es: el 80 por ciento de ganancias viene del 20 por ciento de clientes. A través de un análisis de rentabilidad de clientes, otros han encontrado que los clientes fieles no son necesariamente rentables. Existe una necesidad de comprender el valor que presentan los clientes, valor potencial a largo plazo, y los clientes potenciales que pueden generar beneficios a una institución financiera. Tanto la falta de consideración en las necesidades del cliente como de comprensión que todos los clientes no pueden ser tratados de la misma manera sólo puede conducir a costosos errores de inversión [Peppard, 2000].

De manera general, según [Momparler Pechun, 2008] la situación actual del sector financiero se ha visto caracterizada en los últimos años por varios factores:

- Incremento de la competencia.
- Desarrollo de innovaciones tecnológicas.
- Mayor accesibilidad a los servicios.
- Interacción social de los clientes financieros.
- Exigencia de un menor costo por las transacciones bancarias.
- Clientes de mayor conocimiento financiero.
- Facilidad de acceso a las nuevas tecnologías.

[1] Cross-selling: es la venta de productos y servicios adicionales a un cliente existente. Up-selling: es vender productos y servicios de mayor precio o mayor margen a un cliente existente, [Buttle, 2009]

Esta situación ha obligado a las entidades financieras a incrementar y personalizar su nivel de servicios, así como a involucrar a toda la organización en la búsqueda de una mayor vinculación y rentabilidad de los clientes con el uso de herramientas como el CRM.
Las estrategias tradicionales como reducción de costos y aumentos de precios podrían no ser las mejores opciones. Si bien estas estrategias pueden generar ganancias a corto plazo, en realidad pueden socavar los beneficios a largo plazo. La razón es que los beneficios sostenibles dependen del crecimiento de los saldos. Y ni el recorte directo de los costos ni el aumento de los precios facilitan el crecimiento del saldo, de hecho, ambos pueden disminuirlo. Reducir los costos a menudo conduce a cerrar sucursales y bajar los niveles de servicio, lo cual puede ocasionar que algunos clientes depositen su dinero en otra parte.

En el dominio bancario, definimos ***un cliente desertor como aquel que cierra todas sus cuentas y deja de hacer negocios con el banco.*** Hay muchas razones para que un cliente cierre la (s) cuenta (s). El problema aquí es que el banco no siempre captura este tipo de datos de retroalimentación, por lo tanto, no se puede hacer mayor análisis por lo que este tipo de comportamientos que llevan al abandono de un clientes no puede ser detenido [Chitra and Subashini, 2011].

La importancia de retener un cliente y establecer una relación a largo plazo se debe a que trae los siguientes beneficios para una empresa según [Reichheld and Kenny, 1990]:

1. Primero, los saldos crecen con el tiempo a medida que aumentan los intereses, las cuentas múltiples se consolidan y las posiciones económicas de los clientes mejoran.

2. En segundo lugar, el gasto de adquirir un nuevo cliente se incurre sólo en el primer año. Así que cuanto más larga sea la relación, menor será el costo amortizado de la adquisición de la cuenta.

3. En tercer lugar, el costo de mantener un cliente es relativamente fijo, por lo que los gastos de mantenimiento de cuenta van disminuyendo de los ingresos a medida que la relación se alarga y produce ese nivel esperado más alto de los saldos.

4. Los clientes a largo plazo son más probables en ampliar sus relaciones con otros productos.

5. Los clientes a largo plazo tienen mayor probabilidad de dar buena referencia a sus amigos y parientes al banco.

Una entidad bancaria con alta retención tiene una ventaja a la que los competidores tienen dificultades para reaccionar. Dado que no es fácil rastrear el rendimiento de la retención, es probable que los competidores ni siquiera se den cuenta de que el crecimiento de un banco es impulsado por una mejor retención.
Incluso si los competidores descubren la ventaja de la retención de un banco,

generalmente es muy poco lo que pueden hacer para atraer a sus clientes satisfechos y leales. De hecho, la única manera de competir con un banco de retención alta es construir su propia retención con el fin de limitar el número de clientes que quedan en el mercado.
La retención del cliente también proporciona otras dos ventajas sutiles: un aumento paralelo de la retención de los empleados y la capacidad de añadir el valor a fusiones y adquisiciones [2].
A pesar de que puede parecer trivial, clientes satisfechos significan empleados satisfechos. Es más fácil para los empleados tratar con los clientes que están satisfechos con sus relaciones bancarias que tratar con aquellos que están insatisfechos. Y, por supuesto, los empleados satisfechos tienden a ser empleados de servicio prolongado.
A su vez, la capacidad de retener a los empleados por períodos más largos confiere múltiples beneficios. Una razón obvia es la reducción de los costos de contratación. Por último, los empleados de más tiempo suelen ser capaces de formar relaciones personales con los clientes del banco, entender sus necesidades. Por lo tanto, el aumento de la retención de los clientes lleva a una mayor retención de los empleados, lo que se traduce en una mayor longevidad del cliente.

A pesar de los beneficios de la retención de clientes que parece obvio para muchos, existe una tentación por los bancos en asumir que ya se ha optimizado su retención. Después de todo, podrían pensar que, una cierta cantidad de pérdida del depositante es inevitable.

Para encontrar respuestas a las preguntas qué y por qué es probable que se produzca la deserción, es necesaria una clasificación de los clientes. Para que la predicción de la deserción sea verdaderamente significativa, las causas de la misma deben ser comprendidas. Medir la exactitud de la predicción es otro paso importante abordado en este trabajo [Chitra and Subashini, 2011].

Con el fin de predecir la deserción de los clientes en el sector bancario, se ha hecho uso métodos de minería de datos lo cual implica un proceso iterativo que combina negocios conocimientos, métodos y herramientas de aprendizaje de máquina, además grandes cantidades de información precisa y pertinente para habilitar el descubrimiento de información oculta no-intuitiva en los datos de la empresa.
En el área financiera, la minería de datos se ha aplicado con éxito en la determinación de:

- ¿Quiénes son los probables desertores en los próximos dos meses?
- ¿Quién probablemente es un cliente rentable?
- ¿Cuál es el comportamiento económico de los clientes rentables?

[2] Una fusión es un acuerdo para unir dos compañías existentes en una empresa nueva. Hay varios tipos de fusiones y también varias razones por qué las empresas realizar fusiones. Mayoría de fusiones unen dos empresas existentes en una empresa recién nombrada. Fusiones y adquisiciones se realizan comúnmente para ampliar el alcance de la empresa, expandirse hacia nuevos segmentos o ganar cuota de mercado. Todos estos se hacen para complacer a accionistas y crear valor.

- ¿Qué productos son capaces de comprar los diferentes segmentos?
- ¿Qué propuestas de valor sirven a diferentes grupos?
- ¿Qué atributos caracterizan sus diferentes segmentos?, ¿y cómo cada uno se relaciona con perfil de la persona?

Con el limitado presupuesto de marketing y el personal, los modelos de minería de datos se utilizan para clasificar los clientes y solamente cierto porcentaje de los clientes son contactados por correo electrónico, teléfono, u otro medio con el fin de ofrecerle algún servicio especial. Si el modelo de minería de datos es lo suficientemente bueno y los objetivos están bien definidos, la empresa puede ponerse en contacto con un grupo pequeño de personas con una alta concentración de desertores potenciales [Hu, 2005].

Para predecir clientes propensos a desertar, se han presentado dos grupos de métodos detallados brevemente a continuación.
El primero incluye los métodos de clasificación como: árbol de decisión, regresión logística, bayesiano ingenuo y agrupamiento. Estos métodos pueden ser usados para analizar datos cualitativos y datos cuantitativos, e interpretar los modelos de predicción; sin embargo, estos no pueden garantizar la precisión y la capacidad de generalización de los modelos construidos, para gran escala, alta dimensionalidad, no linealidad, datos de serie de tiempo, etc. El segundo grupo se compone de los métodos de inteligencia artificial, incluyendo: Redes Neuronales Artificiales (ANN: Artificial Neural Network), Mapas de Organización Propia (SOM: Self Organizing Maps), Aprendizaje Evolutivo (EL: Evolutionary Leaning), etc.

Los métodos del segundo grupo pueden superar las dificultades concernientes a los métodos del primer grupo, pues tienen capacidad de mapeo no lineal, robustez y precisión en la predicción. Sin embargo, los métodos basados en la minimización de riesgo empírico siempre conducen a una baja habilidad de generalización y confusa generalización en los modelos [Xie and Ying, 2009]. Para resolver los problemas anteriores se propone usar Máquinas de Soporte Vectorial (SVM, Support Vector Machine) los cuáles, en un contexto de clasificación binaria, intentarán encontrar un hiperplano óptimo lineal tal que el margen de separación entre los dos grupos que se va a separar se maximice. SVM está basado en el principio de la Minimización de Riesgo Estructural (SRM: Structural Risk Minimization), que quiere decir que este tipo de clasificador reduce al mínimo la cota superior sobre el riesgo real, comparado a otros clasificadores que reducen al mínimo el riesgo empírico. Esto causa un funcionamiento de generalización muy bueno [Coussement and Van den Poel, 2008].

En este trabajo se comprobará esta afirmación y se compararán los resultados obtenidos entre regresión logística y SVM para escoger el de mayor exactitud.

El enfoque de la deserción es determinar los clientes que están en riesgo de salir y si es posible en el análisis ver si esos clientes valen la pena la retención. El análisis de deserción depende en gran medida de la definición del deserción del cliente, el

sector empresarial y la relación con los clientes pues influyen en el resultado de cómo se detectan los desertores [Mutanen and Nousiainen, 2006].
Por esta razón, para predecir la deserción, en el capítulo 4 se considerarán para el modelamiento cuestiones como el tipo de industria, en este caso el sector bancario, el tipo de deserción que se predecirá y cómo se definirá la deserción de acuerdo a las necesidades particulares de una entidad financiera. Además se definirá matemáticamente el CLV.

Capítulo 2

Aprendizaje automático

2.1. Preliminares

En vista que uno de los objetivos de este trabajo es predecir la deserción de un cliente, en este capítulo se aborda cuestiones de aprendizaje automático, en particular máquinas de soporte vectorial. Además se describen las etapas que se seguirán en el modelamiento como selección de variables y evaluación.
Una vez que se hayan identificado a cliente rentables, la sección de aprendizaje no supervisado se podría usar para crear segmentos de clientes para establecer estrategias de acuerdo a cada segmento que se haya identificado.

La minería de datos consiste en resolver problemas analizando las bases de datos. Supongamos, que el problema es la fidelidad del cliente inconstante en un mercado altamente competitivo. Una base de datos que contiene las opciones de los clientes y perfiles del mismo contiene la clave para resolver este problema. Los patrones de comportamiento de los antiguos clientes pueden ser analizados para identificar las características distintivas de aquellos que pueden cambiar de producto y los que probablemente permanecerán leales. Una vez que tales características se encuentran, se pueden poner a trabajar los datos para identificar a los clientes actuales que son propensos a pasarse a la competencia.
En la actual economía altamente competitiva, centrada en el cliente y orientada al servicio, los datos son la materia prima que alimenta el crecimiento del negocio, siempre y cuando sea extraída y analizada [Sharp, 2002].

2.2. Minería de datos

Según [Esposito, 2005] el descubrimiento del conocimiento en bases de datos (KDD: Knowledge Discovery from Databases), también llamado Data Mining, combina técnicas de bases de datos, estadísticas y aprendizaje automático. Una aplicación está en la gestión de relaciones con los clientes, específicamente en la deserción de clientes, si los mismos abandonan sus contratos, el beneficio no aumenta. Por lo tanto, la acción de detectar clientes que están a punto de desertar y tomar alguna iniciativa para mantenerlos es una meta importante. La tarea es la

clasificación de los clientes en aquellos que son propensos a cerrar su contrato en un determinado tiempo, y los que continuarán con la empresa. Otras aplicaciones de KDD suelen tener los siguientes objetivos:

- Gestionar la relación con los mejores clientes (CRM), determinar qué productos se ofrecerán (marketing directo o adquisición de clientes) y qué clientes no pagarán (detección de fraude).
- Clasificar los objetos según los perfiles de usuario en los sistemas de recomendación. Los objetos pueden ser, por ejemplo, productos en la tienda de internet como Amazon, o documentos en los motores de búsqueda. En este caso las aplicaciones de KDD no asumen perfiles de usuario, sino que aprenden de las observaciones del comportamiento del usuario.

De acuerdo a [Bhambri, 2013], la minería de datos implica el uso de sofisticadas herramientas de análisis de datos para descubrir patrones y relaciones previamente desconocidos y válidos en grandes conjuntos de datos. Puesto que las herramientas pueden incluir modelos estadísticos, algoritmos matemáticos y métodos de aprendizaje automático, la minería de datos no es sólo la recopilación y gestión de datos, también incluye análisis y predicción. Los datos recopilados o almacenados en la base de datos de la organización necesitan ser procesados para que puedan ser convertidos en información útil y, por lo tanto, puedan usarse para la toma de decisiones. Este proceso de descubrir el conocimiento oculto a partir de datos sin procesar se llama proceso de descubrimiento de conocimiento.

Cuando se determinan claramente las tareas que cumplen con el objetivo de una aplicación, se inicia el proceso KDD. En este trabajo se tratará de aplicar marketing directo una vez que se haya clasificado a los clientes en desertores y no desertores. Además de ver su valor para la empresa y determinar si vale la pena esforzarse en retenerlos.

2.2.1. Minería de datos en el marco de CRM

La minería de datos puede proporcionar una visión del cliente que es vital para establecer una estrategia de CRM eficaz. Puede conducir a interacciones personalizadas con los clientes y, por lo tanto, una mayor satisfacción y relaciones con los clientes rentables a través del análisis de datos.

La minería de datos pueden ayudar en otras tareas a las actividades de marketing como las que se describen a continuación.

Segmentación de clientes: La segmentación es el proceso de dividir la base de clientes en grupos distintos e internamente homogéneos a fin de desarrollar estrategias de marketing diferenciadas según sus características. Existen muchos tipos de segmentación basados en los criterios o atributos específicos utilizados para la segmentación.
En la segmentación conductual, los clientes se agrupan por características de comportamiento y de uso. Aunque los segmentos de comportamiento pueden

crearse con reglas de negocio, este enfoque tiene desventajas inherentes. Puede manejar eficientemente sólo unos pocos campos de segmentación y su objetividad es cuestionable, ya que se basa en las percepciones personales de un experto en negocios. Por otro lado, la minería de datos puede crear segmentos de comportamiento impulsados por datos. Los algoritmos de agrupación pueden analizar datos de comportamiento, identificar los grupos naturales de clientes y sugerir una solución basada en patrones de datos observados. Siempre que los modelos de minería de datos estén debidamente construidos, pueden descubrir grupos con perfiles y características diferentes y conducir a esquemas de segmentación ricos con significado y valor comercial.

Campañas de marketing directo: Los vendedores utilizan campañas de marketing directo para comunicar un mensaje a sus clientes a través de Internet, correo electrónico, telemercadeo (teléfono) y otros canales directos con el fin de evitar la deserción e impulsar la adquisición de clientes y la compra de productos. Cuando no se refinan estas campañas, aunque sean potencialmente eficaces, también pueden conducir a un enorme desperdicio de recursos, a bombardear y molestar a los clientes con comunicaciones no solicitadas. Los modelos de minería y clasificación de datos, en particular, pueden apoyar el desarrollo de campañas de marketing específicas. Analizan las características del cliente y reconocen los perfiles de los clientes objetivo. A continuación se identifican nuevos casos con perfiles similares, se les asigna una alta puntuación de propensión hacia un comportamiento particular y se incluyen en las listas de objetivos.
Los siguientes modelos de clasificación se utilizan para optimizar las campañas de marketing:

1. **Modelos de adquisición:** Estos pueden ser utilizados para reconocer clientes potencialmente rentables mediante la búsqueda de “clones” de valiosos clientes existentes en listas externas de contactos.

2. **Modelos cross / deep / up-selling:** Estos pueden revelar el potencial de compra de los clientes existentes.

3. **Desgaste voluntario o modelos de deserción voluntaria:** Estos identifican señales de deserción tempranas y localizan a aquellos clientes con mayor probabilidad de abandonar voluntariamente.

Cuando están bien construidos, estos modelos pueden identificar los clientes adecuados para contactarlos y llevarlos a las listas de las campañas con mayor cantidad de clientes objetivo. Este camino supera las selecciones al azar, así como las predicciones basadas en las reglas de negocio y la intuición personal [Tsiptsis and Chorianopoulos, 2011].

2.2.2. Proceso de descubrimiento del conocimiento

Un proyecto de minería de datos implica más que el modelado. La fase de modelado es sólo una etapa en el proceso de implementación de un proyecto de

minería de datos. Los pasos de importancia crítica que preceden y siguen a la construcción de los modelos tienen un efecto significativo en el éxito del proyecto.
Existen fases básicas en el desarrollo de un proyecto de minería de datos de acuerdo con el modelo de Proceso Estándar de la Industria para la Minería de Datos (CRISP-DM: Cross Industry Standard Process for Data Mining).
Los proyectos de minería de datos no son simples. Comienzan generalmente con altas expectativas pero pueden terminar en fracaso del negocio si el equipo comprometido no es guiado por un marco metodológico claro. El modelo de proceso CRISP-DM muestra los pasos que deben seguirse para implementaciones exitosas de minería de datos. Estos pasos se describen a continuación:

1. **Comprensión del negocio:** El proyecto de minería de datos debe comenzar con una comprensión del objetivo del negocio y una evaluación de la situación actual. Los parámetros del proyecto deben ser considerados, incluyendo recursos y limitaciones. El objetivo del negocio debe traducirse en un objetivo de minería de datos. Deben definirse criterios de éxito y elaborarse un plan de proyecto.

2. **Comprensión de los datos:** Esta fase implica considerar los requisitos de datos para abordar adecuadamente el objetivo definido y una investigación de la disponibilidad de los datos requeridos. Esta fase también incluye la recolección inicial de datos y la exploración con estadísticas de resumen y herramientas de visualización para entender los datos e identificar posibles problemas de disponibilidad y calidad.

3. **Preparación de los datos:** Los datos a utilizar deben ser identificados, seleccionados y preparados para su inclusión en el modelo de minería de datos. Esta fase implica la adquisición, integración y formato de los datos de acuerdo con las necesidades del proyecto. Los datos consolidados se deben transformarse adecuadamente de acuerdo con los requisitos del algoritmo que se va a aplicar. Los campos nuevos, tales como sumas, promedios, proporciones, indicadores, etc., deben derivarse de los campos originales para enriquecer la información del cliente, resumir mejor las características del cliente y, por lo tanto, mejorar el rendimiento de los modelos.

4. **Modelamiento:** Los datos procesados se usan para el entrenamiento del modelo. Los analistas deben seleccionar la técnica de modelado adecuada para el objetivo de negocio en particular. Antes de la formación de los modelos y especialmente en el caso de la predicción de modelos, el conjunto de datos de modelado debe ser particionado para que el rendimiento del modelo se evalúa en un conjunto de datos independiente. Esta fase implica el examen de algoritmos de modelado alternativos y la configuración de parámetros y una comparación de su ajuste y el rendimiento con el fin de encontrar el que produce los mejores resultados. Basándose en una evaluación inicial de los resultados del modelo, los ajustes del modelo pueden ser revisados y ajustados.

5. **Evaluación:** Los modelos generados se evalúan formalmente no sólo en términos de medidas técnicas sino también, más importante, en el contexto

de los criterios de éxito de negocio establecidos en la fase de comprensión de negocio. El equipo del proyecto debe decidir si los resultados de un modelo dado abordan adecuadamente los objetivos iniciales del negocio. Si es así, este modelo está aprobado y preparado para el despliegue.

6. **Despliegue:** Los hallazgos y conclusiones del proyecto se resumen en un informe, pero este no es el final del proyecto. Incluso el mejor modelo resultará ser un fracaso comercial si sus resultados no se implementan e integran en las operaciones de marketing cotidianas de la organización. Debe diseñarse y desarrollarse un procedimiento que permita la puntuación de los clientes y la actualización de los resultados. El procedimiento de despliegue también debe permitir la distribución de los resultados del modelo en toda la empresa y su incorporación en las bases de datos de la organización y en el sistema operativo CRM. Por último, debe diseñarse un plan de mantenimiento y revisar todo el proceso. Las lecciones aprendidas deben ser tenidas en cuenta y los próximos pasos deben ser planificados.

Las fases anteriores presentan fuertes dependencias y los resultados de una fase pueden conducir a revisar los resultados de las fases precedentes. La naturaleza del proceso es cíclica ya que la minería de datos en sí es un viaje sin fin y búsqueda, exigiendo una continua reevaluación y actualización de las tareas completadas en el contexto de un entorno empresarial en rápido cambio [Tsiptsis and Chorianopoulos, 2011].
En el sistema bancario Indú, los bancos han organizado sus datos de acuerdo con los requisitos para la predicción de la deserción, por lo que se puso de relieve convertir datos crudos de los clientes en datos significativos que se adapten a la modelación del comportamiento de deserción. Entonces, estos datos significativos pueden convertirse en conocimiento utilizando técnicas predictivas de minería de datos las cuales ayudan a las empresas en particular a la banca, las telecomunicaciones, los seguros y la venta minorista a construir un perfil del cliente preciso basado en el comportamiento del mismo.
La minería de datos puede ayudar a dirigirse a clientes nuevos y ofrecer productos y servicios descubriendo los patrones de compra anteriores del cliente a fin de que el banco pueda retener a los clientes existentes ofreciendo incentivos que se adaptan individualmente a las necesidades de cada cliente, como se muestra en los siguientes ejemplos:

- Cuando el Chase Manhattan Bank en Nueva York empezó a perder clientes frente a sus competidores, comenzó a utilizar minería de datos para analizar las cuentas de los clientes e hizo cambios en los requisitos de su cuenta, permitiendo así al banco retener a sus clientes rentables.

- La minería de datos también está siendo utilizada por Fleet Bank Boston para identificar a los mejores candidatos para ofertas de fondos mutuos. El banco analiza los datos demográficos de los clientes y los datos de las cuentas a lo largo de diferentes líneas de productos para determinar qué clientes pueden

invertir en un fondo mutuo y esta información se utiliza para orientar a esos clientes.

- El servicio de atención al cliente de America West Coast tiene a sus representantes listos con los perfiles de los clientes resumidos por la minería de datos para lanzar nuevos productos y servicios que son los más relevantes para cada cliente.

Las técnicas de minería de datos pueden ser de gran ayuda para la organización en la solución de problemas de negocios mediante la búsqueda de patrones, asociaciones y correlaciones que se oculta en la información comercial almacenada en las bases de datos. Recordemos que las organizaciones pueden utilizar técnicas de KDD para la adquisición de nuevos clientes, la detección de fraude, la prestación de productos basados en el segmento para una mejor orientación de los clientes, el análisis de los patrones de compra de los clientes que con el tiempo ayuda a una mejor retención, relación, y por lo tanto todas estas actividades puede ser información útil para predecir el comportamiento de abandono del cliente antes de tiempo.
Las organizaciones normalmente enfrentan varios problemas de implementación, los cuales deben ser abordados antes de ir a la ejecución final. Sin embargo, mediante la selección de las técnicas de minería de datos y su correcta aplicación puede resultar una bendición para la organización que será capaz de evitar la deserción de los clientes y llevará a una relación a largo plazo [Bhambri, 2013].

2.2.3. Modelos Predictivos / Supervisados y técnicas

Según [Tsiptsis and Chorianopoulos, 2011], en el modelo supervisado o predictivo, directo o dirigido, el objetivo es predecir un evento o estimar los valores de un atributo numérico discreto o continuo. En estos modelos hay campos o atributos de entrada y un campo de salida o de destino. Los campos de entrada también se denominan predictores porque son utilizados por el modelo para identificar una función de predicción para el campo de salida. Podemos pensar en los predictores como la parte X de la función y el campo objetivo como la parte Y, el resultado.
El modelo utiliza los campos de entrada que se analizan con respecto a su efecto en el campo de destino. El reconocimiento de patrones es “supervisado” por el campo de destino. Las relaciones se establecen entre los campos de entrada y salida. El modelo genera una “función” de asignación de entrada-salida, que asocia predictores con la salida y permite la predicción de los valores de salida, dados los valores de los campos de entrada. Los modelos predictivos se clasifican además en modelos de clasificación y estimación:

- **Modelos de clasificación o propensión:** En estos modelos los grupos o clases objetivo se conocen desde el principio. El objetivo es clasificar los casos en estos grupos predefinidos, en otras palabras, predecir un evento. El modelo generado puede utilizarse como un motor de puntuación para asignar nuevos casos a las clases predefinidas. También estima un puntaje de propensión para cada caso. El puntaje de propensión denota la probabilidad de ocurrencia del grupo o evento objetivo.

- **Modelos de estimación:** Estos modelos son similares a los modelos de clasificación pero con una diferencia importante. Se usan para predecir el valor de un campo continuo basado en los valores observados de los atributos de entrada.

En el modelado supervisado, ya sea para la predicción de un evento o para un resultado numérico continuo, se requiere la disponibilidad de un conjunto de datos de formación con datos históricos. Los modelos aprenden de casos pasados. Para que los modelos predictivos asocien patrones de datos de entrada con resultados específicos, es necesario presentarlos con casos con resultados conocidos. Esta fase se llama fase de entrenamiento. Durante esa fase, el algoritmo predictivo construye la función que conecta las entradas con el campo de destino. Una vez que las relaciones se identifican y el modelo es evaluado y con poder predictivo satisfactorio, la fase de puntuación sigue. Los nuevos registros, para los cuales los valores de resultado son desconocidos, se presentan al modelo y se anotan en consecuencia.
Algunos modelos predictivos como la regresión y los árboles de decisión son transparentes, proporcionando una explicación de sus resultados. Además de la predicción, estos modelos también se pueden utilizar para la percepción y caracterización de los clientes. Pueden identificar entradas con un efecto significativo en el atributo de destino y pueden revelar el tipo y la magnitud del efecto. Por ejemplo, los modelos supervisados se pueden aplicar para encontrar los controladores asociados con la satisfacción del cliente o deserción. Del mismo modo, los modelos supervisados también pueden complementar las técnicas de informes tradicionales en el perfil de los segmentos de una organización mediante la identificación de las características diferenciadoras de cada grupo.
Según el nivel de medición del campo a predecir, los modelos supervisados se clasifican además en:

- Técnicas de modelización de clasificación o propensión.

- Técnicas de modelización de estimación o regresión.

Cuando el objetivo es categórico, el uso de un modelo de clasificación es apropiado. Estos modelos analizan resultados discretos y se utilizan para clasificar nuevos registros en las clases predefinidas. En otras palabras, predicen eventos. Las puntuaciones de confianza complementan sus predicciones, denominada la probabilidad de un resultado particular. Cuando se desea estimar valores de resultado continuo, se aplican modelos de estimación. Estos modelos son también contemplados como modelos de regresión.
Otro uso de modelos supervisados está en la selección de predictores. Estos modelos se utilizan como un paso preparatorio antes del desarrollo de un modelo predictivo. Evalúan la importancia predictiva de los campos de entrada originales e identifican los predictores significativos. Los predictores con poca o ninguna potencia predictiva se eliminan de los siguientes pasos de modelado.

Encontrar predictores útiles con modelos supervisados de exploración de campo

Otra clase de técnicas de modelado supervisado incluye los modelos supervisados de exploración de campo. Estos son modelos que usualmente sirven como un paso de preparación para el desarrollo de modelos de clasificación y estimación. La situación de tener cientos o incluso miles de predictores candidatos no es inusual en complicadas tareas de minería de datos. Sin embargo, algunos de estos campos pueden no influir en el campo de salida que queremos predecir. El papel de los modelos supervisados de exploración de campo es evaluar todos los insumos disponibles y encontrar los predictores clave y los predictores con poca o ninguna importancia que son candidatos para la eliminación potencial del modelo predictivo.
Algunos algoritmos predictivos, incluyendo árboles de decisión, por ejemplo, integran mecanismos de exploración que filtran internamente los predictores no relacionados. Los modelos de exploración de campo pueden reducir eficientemente la dimensionalidad de los datos conservando sólo aquellos campos relevantes para el resultado de interés, [Tsiptsis and Chorianopoulos, 2011].
Los modelos de cribado de campo se usan generalmente en la fase de preparación de datos de un proyecto de minería de datos con el fin de realizar las siguientes tareas:

- Evaluar la calidad de los predictores potenciales. Incorporan criterios específicos para identificar predictores inadecuados: por ejemplo, predictores con un gran porcentaje de valores faltantes (nulos), predictores continuos constantes o con poca variación, predictores categóricos con demasiadas categorías o casi todos los registros pertenecientes a una sola categoría.

- Clasificar los predictores de acuerdo con su poder predictivo. Se evalúa la influencia de cada predictor en el campo objetivo y se calcula una medida de importancia. Los predictores se ordenan en consecuencia.

- Filtrar los predictores sin importancia. Se identifican predictores no relacionados con el campo de destino. Los analistas tienen la opción de filtrarlos, reduciendo el conjunto de campos de entrada a los relacionados con el campo de destino.

El propósito de todos los modelos de clasificación es proporcionar información y ayudar en el refinamiento y la optimización de las aplicaciones de marketing. El primer paso después del entrenamiento del modelo es explorar los resultados generados, que pueden venir en diferentes formas según el modelo utilizado: reglas, ecuaciones, gráficos. La extracción del conocimiento es seguida por la evaluación de la eficiencia predictiva del modelo y por el despliegue de los resultados con el fin de clasificar los nuevos registros de acuerdo con los resultados del modelo.

Las siguientes técnicas de modelado están incluidas en la clase de los modelos de clasificación:

- Árboles de decisión

- Regresión logística
- Máquinas de soporte vectorial

En este trabajo se utilizó árboles de decisión para la selección de variables significativas en el modelo de deserción.
Para la clasificación se usó regresión logística y máquinas de soporte vectorial con el fin de comparar resultados como se verá en el capítulo 4.

2.2.4. Evaluación de Modelos de Clasificación

Antes de aplicar el modelo generado en nuevos registros, se requiere un procedimiento de evaluación para evaluar su capacidad predictiva. Se anotan los datos históricos con resultados conocidos, que se utilizaron para la formación del modelo, y se derivan dos nuevos campos: la categoría de resultados previstos y la respectiva puntuación de confianza.
En la práctica, los modelos nunca son tan precisos. Siempre hay errores y registros mal clasificados. Una comparación de los valores previstos con los reales es el primer paso para evaluar el desempeño del modelo.
Esta comparación proporciona una estimación de la precisión predictiva futura del modelo en casos no vistos. Para que este procedimiento sea más válido, es recomendable evaluar el modelo en un conjunto de datos que no se utilizó para la formación del modelo. Esto se logra dividiendo el conjunto de datos históricos en dos partes distintas a través del muestreo aleatorio: el entrenamiento y el conjunto de datos de prueba. Una práctica común es asignar aproximadamente el 70-75 % de los casos al conjunto de datos de entrenamiento. Los procedimientos de evaluación se aplican a ambos conjuntos de datos. Los analistas deben centrarse principalmente en el examen de los indicadores de rendimiento en el conjunto de datos de pruebas. Un modelo de bajo rendimiento en el conjunto de datos de prueba debe ser reexaminado, ya que este es un signo típico de sobre-estimación (overfitting) y de la memorización de los datos de formación específica. Los modelos con este comportamiento no proporcionan resultados generalizables. Proporcionan soluciones que sólo funcionan para los datos particulares en los que fueron entrenados. Algunos analistas usan las pruebas dataset para refinar los parámetros de los modelos y dejar una tercera parte de los datos, a saber la validación dataset para la evaluación. Sin embargo, el mejor acercamiento, que lamentablemente no siempre es empleado, sería probar el funcionamiento del modelo en un tercio de un período de tiempo diferente.
Uno de los indicadores de rendimiento más comunes para los modelos de clasificación es la tasa de error. Mide el porcentaje de errores de clasificación. La tasa de error global indica el porcentaje de registros que no fueron clasificados correctamente por el modelo. Dado que algunos errores pueden ser más costosos que otros, este porcentaje también se estima para cada categoría del campo objetivo. La tasa de error se resume en matrices de clasificación errónea o coincidencia o confusión, [Tsiptsis and Chorianopoulos, 2011].

2.2.5. Modelos no supervisados y técnicas

Según [Tsiptsis and Chorianopoulos, 2011], en el caso de las técnicas de modelización supervisadas, las mismas se usan para clasificación, estimación o selección de campo, su característica común es que todos ellos implican un atributo objetivo que debe estar asociado con un conjunto examinado de entradas. El entrenamiento del modelo y el reconocimiento del patrón de datos son guiados o supervisados por un campo objetivo. Este no es el caso en el modelado sin supervisión donde todos los insumos se tratan por igual con el fin de extraer información que se pueda utilizar, principalmente, para la identificación de agrupaciones y asociaciones.
Al agregar el factor de tiempo al modelado de la asociación tenemos el modelado de secuencia: en el modelado de secuencias se analiza las asociaciones a través del tiempo y tratamos de descubrir la serie de eventos, el orden en que suceden los eventos. A veces cuando sólo interesa identificar registros que “no encajan bien”, es decir, registros con patrones de datos inusuales e inesperados. En tales casos, las técnicas de detección de registros pueden emplearse como un paso de auditoría de datos antes de construir un modelo subsiguiente para detectar registros anormales (anómalos).
Las técnicas de agrupación son muy populares y su uso está muy extendido en la minería de datos y en la investigación de mercado pues pueden apoyar el desarrollo de diferentes esquemas de segmentación de acuerdo con los atributos de agrupamiento utilizados: a saber, la segmentación conductual, actitudinal o demográfica.

En modelos no supervisados o no dirigidos no hay campo de salida, solo entradas. El reconocimiento de patrones no está dirigido no está guiado por un atributo de destino específico. El objetivo de estos modelos es descubrir los patrones de datos en el conjunto de campos de entrada. Los modelos no supervisados incluyen:

- **Modelos de agrupación:** En estos modelos los grupos no se conocen de antemano, en su lugar, los algoritmos analizan los patrones de datos de entrada e identifican los grupos naturales de registros o casos. Cuando los nuevos casos son anotados por el modelo de agrupación generado se asignan a uno de los grupos revelados.

- **Modelos de asociación y secuencia:** Estos modelos también pertenecen a la clase de modelización no supervisada. No implican la predicción directa de un solo campo, de hecho, todos los campos involucrados tienen un doble papel, ya que actúan como entradas y salidas al mismo tiempo. Los modelos de asociación detectan asociaciones entre eventos, productos o atributos discretos. Los modelos de secuencia detectan las asociaciones a lo largo del tiempo.

Las técnicas de agrupación incluyen las siguientes:

- Aglomerativo o jerárquico

- K-medias

- Clúster de dos etapas
- Redes Kohonen/Mapas auto-organizados (SOM: Self-Organizing Map)

2.3. Machine Learning

Las técnicas de análisis de datos que se han utilizado tradicionalmente para tales tareas incluyen análisis de regresión, análisis de conglomerados, taxonomía numérica, análisis multidimensional, otros métodos estadísticos multivariados, modelos estocásticos, análisis de series de tiempo, técnicas de estimación no lineal y otros. Estas técnicas han sido ampliamente utilizadas para resolver muchos problemas prácticos. Sin embargo, se orientan principalmente hacia la extracción de datos cuantitativos y estadísticos y, como tales, tienen limitaciones inherentes. Por ejemplo, un análisis estadístico puede determinar covarianzas y correlaciones entre variables en los datos. Sin embargo, no pueden caracterizar las dependencias en un nivel y procedimiento abstractos, conceptuales, una explicación casual de las razones por las cuales estas dependencias existen. Tampoco pueden desarrollar una justificación de estas relaciones en forma de descripciones y leyes de estilo lógico de nivel superior. Un análisis de datos estadísticos puede determinar la tendencia central y la varianza de los factores dados, y un análisis de regresión puede encajar una curva a un conjunto de puntos de datos. Estas técnicas no pueden, sin embargo, producir una descripción cualitativa de las regularidades y determinar su dependencia de factores no explícitamente proporcionados en los datos, ni pueden hacer una analogía entre la regularidad descubierta y la regularidad en otro dominio. Una técnica de taxonomía numérica puede crear una clasificación de entidades y especificar una similitud numérica entre las entidades ensambladas en la misma o en diferentes categorías. Sin embargo, no puede construir una descripción cualitativa de las clases creadas y razones de hipótesis para que las entidades estén en la misma categoría. Los atributos que definen la similitud, así como las medidas de similitud, deben ser definidos previamente por un analista de datos. Además, estas técnicas no pueden por sí mismas basarse en conocimientos de dominio de fondo para generar automáticamente atributos relevantes y determinar su relevancia cambiante para diferentes problemas de análisis de datos.

Para abordar tareas como las enumeradas anteriormente, un sistema de análisis de datos debe estar equipado con una cantidad sustancial de antecedentes y ser capaz de realizar tareas de razonamiento simbólico que involucren ese conocimiento y los datos. En resumen, las técnicas tradicionales de análisis de datos facilitan interpretaciones de datos útiles y pueden ayudar a generar información importante sobre los procesos detrás de los datos. Estas interpretaciones y percepciones son el último conocimiento buscado por aquellos que construyen bases de datos. Sin embargo, este conocimiento no es creado por estas herramientas, sino que debe ser derivado por el análisis de datos humanos.

En los esfuerzos por satisfacer la creciente necesidad de nuevas herramientas de análisis de datos que superen las limitaciones anteriores, los investigadores han

recurrido a ideas y métodos desarrollados en el aprendizaje automático. El campo del aprendizaje automático es una fuente natural de ideas para este propósito, ya que la esencia de la investigación en este campo es desarrollar modelos computacionales para adquirir conocimiento a partir de hechos y conocimientos de fondo. Estos y otros esfuerzos relacionados han llevado al surgimiento de un nuevo área de investigación, frecuentemente llamada minería de datos y descubrimiento de conocimiento.

La estadística como el campo tradicional de la inferencia ha proporcionado modelos con suposiciones más o menos detalladas sobre la distribución de datos. La teoría clásica de la inferencia bayesiana ha demostrado su utilidad en una gran variedad de dominios de aplicación que van desde aplicaciones médicas hasta datos de consumo y análisis de canasta de mercado. Además de los métodos clásicos, las redes neuronales y el aprendizaje automático han aportado ideas, conceptos y algoritmos al análisis de estos conjuntos de datos con un nuevo sabor distintivo. Los nuevos enfoques planteados por estos investigadores en la última década se apartan del análisis tradicional de datos estadísticos de varias maneras: dependen menos de los supuestos estadísticos sobre la distribución real de los datos, dependen menos de los modelos que permiten un análisis matemático simple, son modelos que pueden aprender complicadas dependencias no lineales de grandes conjuntos de datos. Considerando que la estadística ha sido durante mucho tiempo una herramienta para probar las teorías propuestas por los científicos, el aprendizaje de máquinas y la investigación de redes neuronales son más bien evaluados sobre la base de lo bien que generalizar en nuevos datos, que provienen del mismo proceso desconocido que generó los datos de formación.

Los usuarios de muchos servicios de información comerciales, gubernamentales y privados se beneficiarán de nuevas tecnologías de aprendizaje automático que explotan nuevos conocimientos integrando y analizando grandes cantidades de datos ampliamente distribuidos para descubrir e informar sobre relaciones sutiles y patrones de eventos que no son inmediatamente perceptibles por Inspección humana directa [Tsiptsis and Chorianopoulos, 2011].
El aprendizaje automático proporciona la base técnica de la minería de datos. Se utiliza para extraer información de los datos en bruto en las bases de datos, información que se expresa en una forma comprensible y puede utilizarse para una variedad de propósitos. El proceso es de abstracción: tomar los datos, con todas sus imperfecciones, e inferir cualquier estructura que lo subyace. Algunas aplicaciones de minería de datos se centran en la predicción: pronosticar lo que sucederá en nuevas situaciones a partir de datos que describen lo que pasó, a menudo adivinando la clasificación de nuevos ejemplos. Pero estamos igualmente interesados en aplicaciones en las que el resultado de “aprender” es una descripción real de una estructura que se puede utilizar para clasificar ejemplos. Esta descripción estructural apoya la explicación, la comprensión y la predicción. Los conocimientos obtenidos por los usuarios de las aplicaciones son de mayor interés en la mayoría de las aplicaciones prácticas de minería de datos; de hecho, ésta es una de las principales ventajas del aprendizaje automático sobre el modelado estadístico clásico [Witten and Pal, 2016].

El objetivo del aprendizaje automático es diseñar algoritmos de aprendizaje que realicen el aprendizaje automáticamente sin intervención ni asistencia humana. En lugar de programar la computadora para resolver la tarea directamente, en el aprendizaje de la máquina, se busca métodos por los cuales la computadora presentará su propio programa basado en ejemplos que proporcionamos.
Se dice que un programa de computadora aprende de la experiencia E con respecto a alguna clase de tareas T y medida de desempeño P, si su desempeño en tareas en T, medido por P, mejora con la experiencia E.
Un sistema de aprendizaje está caracterizado por los siguientes elementos:

1. tarea o tareas, T
2. experiencia, E
3. medida de desempeño, P.

Así, un sistema de aprendizaje genérico puede ser definido por los siguientes componentes:

- *Objetivo*: Definido con respecto a la tarea a realizar por el sistema;
- *Modelo*: Una función matemática que mapea la percepción a las acciones;
- *Reglas de aprendizaje*: Se utiliza para actualizar los parámetros del modelo con una nueva experiencia de una manera que optimiza las medidas de rendimiento con respecto a las metas. Las reglas de aprendizaje ayudan al algoritmo a buscar el mejor modelo;
- *Experiencia*: Un conjunto de percepciones (y posiblemente las acciones correspondientes).

Antes de entrar en el núcleo de un sistema de aprendizaje, primero tenemos que definir las nociones básicas.
Un *ejemplo* (a veces también llamado una *instancia*) es el objeto que se está clasificando. Por ejemplo, si consideramos la clasificación de los pacientes con tumores de cáncer, entonces los pacientes son los ejemplos.
Un ejemplo es descrito por un conjunto de *atributos*, también conocidos como *características* o *variables*.
Por ejemplo, la clasificación de pacientes con tumores cancerosos, un paciente podría describirse por atributos tales como sexo, edad, peso, tamaño del tumor, forma del tumor, presión arterial, etc.
La *etiqueta* es la categoría que estamos tratando de predecir. Por ejemplo, en la clasificación de pacientes con cáncer, las etiquetas pueden ser “avascular”, “vascular” y “angiogénesis”.
Durante el entrenamiento, el algoritmo de aprendizaje se suministra con *ejemplos etiquetados*, mientras que durante la prueba, sólo se proporcionan *ejemplos no etiquetados*.
En cierta situación podemos suponer que sólo dos etiquetas son posibles que también podríamos llamar 0 y 1 (por ejemplo, cáncer o no cáncer). Esto hará las

cosas muy simples.
También haremos la suposición simplificadora de que existe un mapeo de ejemplos a etiquetas. Este mapeo se llama un *concepto*. Así, un concepto es una función de la forma:
$c : X \rightarrow \{0, 1\}$
Donde X es el espacio de todos los ejemplos posibles llamados el *dominio* o el *espacio de la instancia*. Una colección de conceptos se llama una *clase de concepto*. A menudo se supone que los ejemplos han sido etiquetados por un concepto desconocido de una clase de concepto *conocido* [Grosan and Abraham, 2011].

2.3.1. Pasos a seguir en el aprendizaje

Los pasos principales en un proceso de aprendizaje son los siguientes:

- *Datos y supuestos*: Los datos se refieren a los datos disponibles para la tarea de aprendizaje y los supuestos representan lo que podemos asumir sobre el problema.
- *Representación*: Debemos definir cómo representar los ejemplos a clasificar. Hay muchos métodos de representación para los mismos datos. La elección de la representación puede determinar si la tarea de aprendizaje es muy fácil o muy difícil.
- *Método y estimación*: El método tiene en cuenta cuáles son las posibles hipótesis y la estimación ayuda a ajustar las predicciones basadas en la retroalimentación (como actualizar los parámetros cuando hay un error, etc.).
- *Evaluación*: Esto evalúa el funcionamiento del método (por ejemplo, considerando la proporción de datos clasificados erróneos y el conjunto de datos).
- *Selección del modelo*: Nos dice si podemos considerar el enfoque para hacerlo aún mejor o para hacerlo más flexible, o, si deberíamos elegir un modelo completamente diferente que sería más adecuado.

[Grosan and Abraham, 2011]

En este trabajo se desarrollará un aprendizaje supervisado utilizando máquinas de soporte vectorial cuyo objetivo es clasificar un grupo de clientes en desertores y no desertores.

Capítulo 3

MÁQUINAS DE SOPORTE VECTORIAL (SUPPORT VECTOR MACHINE, SVM)

Este capítulo se centra en la teoría de separación de un conjunto de datos en dos clases mediante máquinas de soporte vectorial y considerando el desbalance entre la clase desertora y no desertora.

3.1. Minimización del Riesgo Empírico

Mientras la mayoría de los métodos de aprendizaje se centran en minimizar los errores cometidos por el modelo generado a partir de los ejemplos de entrenamiento (error empírico), el sesgo inductivo asociado a SVM radica en la minimización del denominado *riesgo estructural*.

Los fundamentos de las máquinas de soporte vectorial han ganado popularidad debido a muchas características prometedoras, como un mejor rendimiento empírico. La formulación utiliza el principio de minimización de riesgos estructurales (SRM), que ha demostrado ser superior al principio de minimización de riesgos empíricos (ERM) tradicional. SRM minimiza un límite superior en el riesgo esperado, mientras que ERM minimiza el error en los datos de entrenamiento. Esta diferencia es la que equipa a SVM con una mayor capacidad de generalización, que es el objetivo del aprendizaje estadístico.
La capacidad de una hipótesis para clasificar correctamente los datos que no están en el conjunto de entrenamiento se conoce como su generalización, [Jakkula, 2006].

En el caso del reconocimiento de patrones de dos clases, la tarea de aprender a partir de ejemplos se puede formular de la siguiente manera, dado un conjunto de funciones de decisión:

$$\{f_w(\mathbf{x}) : w \in \Omega\}, \qquad f_w : \mathbb{R}^n \to \{-1, 1\},$$

donde Ω es un conjunto de parámetros, y un conjunto de ejemplos

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_l, y_l), \qquad \mathbf{x}_i \in \mathbb{R}^n, y_i \in \{-1, 1\}$$

extraídos de una distribución desconocida $P(\mathbf{x}, y)$, queremos encontrar una función f_{w^*} que proporciona el menor valor posible para el *riesgo esperado*:

$$R(w) = \int |f_w(\mathbf{x}) - y| P(\mathbf{x}, y) d_{\mathbf{x}} d_y.$$

Las funciones f_w son usualmente llamadas *hipótesis*, y el conjunto $\{f_w(\mathbf{x}) : w \in \Omega\}$ es llamado el espacio de hipótesis denotado por $\mathscr{H}$. El riesgo esperado es por lo tanto una medida de cuán buena es una hipótesis para predecir la etiqueta correcta y para un punto $\mathbf{x}$.
Dado que la distribución de probabilidad $P(\mathbf{x}, y)$ es desconocida no podemos calcular, y por lo tanto minimizar, el riesgo esperado $R(w)$. Sin embargo, dado que tenemos acceso a un muestreo de $P(\mathbf{x}, y)$, podemos calcular una aproximación estocástica de $R(w)$, el llamado riesgo empírico:

$$R_{emp}(w) = \frac{1}{l} \sum_{i=1}^{l} |f_w(\mathbf{x}_i) - y|.$$

Dado que la ley de las grandes números garantiza que el riesgo empírico converje en probabilidad al riesgo esperado, un enfoque común consiste en minimizar el riesgo empírico más que el riesgo esperado. La intuición que subyace a este enfoque (*Principio de minimización de riesgos empíricos*) es que si R_{emp} converge a R, el mínimo de R_{emp} puede converger al mínimo de R. Si la convergencia del mínimo de R_{emp} al mínimo de R no se cumple, el principio de minimización de riesgos empíricos no permite hacer inferencias basadas en el conjunto de datos, por lo que se dice que no es consistente. Como lo muestran Vapnik y Chervonenkis, la consistencia tiene lugar si y solo si la convergencia en la probabilidad de R_{emp} a R se reemplaza por una convergencia uniforme en la probabilidad. Vapnik y Chervonenkis demostraron que la condición necesaria y suficiente para la consistencia del principio de minimización de riesgos empíricos es la finitud de la dimensión-VC h del espacio de hipótesis $\mathscr{H}$. La dimensión VC del espacio de hipótesis $\mathscr{H}$ (o dimensión VC del clasificador f_w) es un número natural, posiblemente infinito, que es, hablando en forma ligera, el mayor número de puntos de datos que se pueden separar de todas las formas posibles mediante ese conjunto de funciones f_w. La dimensión VC es una medida de la complejidad del conjunto $\mathscr{H}$, y a menudo, pero no necesariamente, es proporcional al número de parámetros libres del clasificador f_w.
La teoría de la convergencia uniforme en probabilidad desarrollada por Vapnik y Chervonenkis también proporciona límites en la desviación del riesgo empírico del riesgo esperado. Un límite uniforme típico de Vapnik y Chervonenkis, que se mantiene con la probabilidad $1 - \eta$, tiene la siguiente forma:

$$R(w) \leq R_{emp}(w) + \sqrt{\frac{h(ln\frac{2l}{h} + 1) - ln\frac{\eta}{4}}{l}}, \qquad \forall w \in \Omega \tag{3.1}$$

donde h es la dimensión-VC de f_w. A partir de este límite, está claro que, para lograr un pequeño riesgo esperado, es decir, buenos resultados de generalización, tanto el riesgo empírico como la relación entre la dimensión VC y el número de puntos

de datos debe ser pequeña. Dado que el riesgo empírico suele ser una función decreciente de h, resulta que, para un número dado de puntos de datos, existe un valor óptimo de la dimensión-VC. La elección de un valor apropiado para h (que en la mayoría de las técnicas se controla mediante el número de parámetros libres del modelo) es crucial para obtener buenos resultados, especialmente cuando el número de puntos de datos es pequeño. El límite en 3.1 sugiere que el principio de minimización de riesgos empíricos puede ser reemplazado por un mejor principio de inducción, como veremos en la próxima sección.

3.2. Minimización de Riesgo Estructural

La técnica de minimización de riesgos estructurales desarrollada por Vapnik es un intento de superar el problema de elegir una dimensión VC apropiada. Está claro de la ecuacion 3.1 que un pequeño valor del riesgo empírico no implica necesariamente un pequeño valor del riesgo esperado. Un principio de inducción diferente, llamado el Principio de Minimización del Riesgo Estructural, ha sido propuesto por Vapnik. El principio se basa en la observación de que, para hacer que el riesgo esperado sea pequeño, ambos lados en la ecuación 3.1 deberían ser pequeños. Por lo tanto, tanto la dimensión VC como el riesgo empírico deberían minimizarse al mismo tiempo. Para implementar el principio de SRM, se necesita una estructura anidada de espacios de hipótesis.

$$\mathscr{H}_1 \subset \mathscr{H}_2 \subset ... \subset \mathscr{H}_n \subset ...$$

con la propiedad que $h(n) \leq h(n+1)$ donde $h(n)$ es la dimension-VC de conjunto $\mathscr{H}$. La ecuación 3.1 sugiere que sin tener en cuenta los factores logarítmicos, el siguiente problema debe ser resuelto de la siguiente manera:

$$\min_{\mathscr{H}_n} \left(R_{emp}[w] + \sqrt{\frac{h(n)}{l}} \right) \tag{3.2}$$

El principio SRM está claramente fundado matemáticamente, pero puede ser difícil de implementar por las siguientes razones:

1. La dimensión-VC de $\mathscr{H}$ podría ser difícil de calcular, y solo hay un pequeño número de modelos para los cuales se conoce cómo calcular la dimensión VC.

2. Aun suponiendo que se pueda calcular la dimensión-VC de $\mathscr{H}$, no es fácil resolver el problema de minimización 3.2. En la mayoría de los casos, se tendrá que minimizar el riesgo empírico para cada conjunto $\mathscr{H}$, y luego elegir el $\mathscr{H}$ que minimice la ecuación 3.2.

Por lo tanto, la implementación de este principio no es fácil, porque no es trivial controlar la dimensión-VC de una técnica de aprendizaje durante la fase de entrenamiento. El algoritmo SVM logra este objetivo, minimizando un límite en la dimensión VC y el número de errores de entrenamiento al mismo tiempo. En la siguiente sección discutiremos esta técnica en detalle y mostraremos cómo su implementación está relacionada con la programación cuadrática, [Osuna et al., 1997].

3.3. Propiedades de las máquinas de soporte vectorial

Conocidos por su robustez, buena capacidad de generalización y soluciones óptimas globales únicas, las Máquinas de Soporte Vectorial son probablemente el enfoque de aprendizaje automático más popular para el aprendizaje supervisado. Desde su introducción en 1992, SVM han encontrado su camino en una gran variedad de aplicaciones, como predicción del clima, clasificación de defectos, reconocimiento de altavoces, identificación de escritura a mano, procesamiento de imágenes y audio, análisis de video y diagnóstico médico.

Lo que hace que SVM sea un atractivo marco de aprendizaje automático se puede resumir en las siguientes propiedades:

- *SVM es una técnica dispersa*. Al igual que los métodos no paramétricos, SVM requiere que todos los datos de entrenamiento estén disponibles, es decir, que estén almacenados en la memoria durante la fase de entrenamiento, cuando se aprenden los parámetros del modelo SVM. Sin embargo, una vez que se identifican los parámetros del modelo, la SVM depende solo de un subconjunto de estas instancias de capacitación, denominadas vectores de soporte, para la predicción futura. Los vectores de soporte definen los márgenes de los hiperplanos. Los vectores de soporte se encuentran después de un paso de optimización que implica una función objetivo regularizada por un término de error y una restricción, utilizando la relajación de Lagrange. La complejidad de la tarea de clasificación con SVM depende del número de vectores de soporte en lugar de la dimensionalidad del espacio de entrada. La cantidad de vectores de soporte que finalmente se conservan del conjunto de datos original depende de los datos y varía según la complejidad de los datos, que se captura mediante la dimensionalidad de los datos y la separabilidad de clase. El límite superior para el número de vectores de soporte es la mitad del tamaño del conjunto de datos de entrenamiento, pero en la práctica esto rara vez es el caso. El modelo SVM descrito matemáticamente en este capítulo está escrito como una suma ponderada de los vectores de soporte, lo que le da al marco SVM las mismas ventajas que las técnicas paramétricas en términos de tiempo computacional reducido para los requisitos de prueba y almacenamiento.

- *SVM es una técnica de kernel*. Cuando los datos son no separables, SVM utiliza un kernel para mapear los datos en un espacio de dimensiones superiores antes de resolver la tarea de aprendizaje automático como un problema de optimización convexa en el que los óptimos se encuentran analíticamente en lugar de heurísticamente, como con otras técnicas de aprendizaje automático. A menudo, los datos de la vida real no son separables linealmente en el espacio de entrada original. En otras palabras, las instancias que tienen etiquetas diferentes comparten el espacio de entrada de una manera que impide que un hiperplano lineal separe correctamente las diferentes clases involucradas en esta tarea de clasificación. SVM mapea los datos, usando funciones de kernel predefinidas, en un espacio nuevo pero

de mayor dimensión, donde un separador lineal podría discriminar entre las diferentes clases. La fase de optimización de SVM implicará aprender solo una superficie discriminante lineal en el espacio mapeado. Por supuesto, la selección y la configuración de la función kernel son cruciales para la optimización de SVM.

- *SVM es un separador de margen máximo.* Más allá de minimizar el error o una función de costo, basado en los conjuntos de datos de entrenamiento, SVM impone una restricción adicional al problema de optimización: el hiperplano debe ubicarse de tal manera que esté a una distancia máxima de las diferentes clases. Tal término fuerza el paso de optimización para encontrar el hiperplano que eventualmente se generalizaría mejor porque está situado a una distancia igual y máxima de las clases. Esto es esencial, porque el entrenamiento se realiza en una muestra de la población, mientras que la predicción se realizará en instancias aún por ver que pueden tener una distribución ligeramente diferente de la del subconjunto capacitado.

- *SVM utiliza la minimización de riesgos estructurales (SRM) y satisface los requisitos de dualidad y convexidad.* SRM es un principio inductivo que selecciona un modelo para el aprendizaje a partir de un conjunto de datos de entrenamiento finito. Como indicador de control de capacidad, SRM propone una compensación entre las dimensiones de VC, es decir, la hipótesis de la complejidad del espacio y el error empírico. La formulación de SRM es una optimización convexa con n variables en la función de costo a maximizar y m restricciones, que se pueden resolver en tiempo polinomial. SRM utiliza un conjunto de modelos secuenciados en un orden creciente de complejidad. Para los modelos no complejos, el error es alto porque un modelo simple no puede capturar toda la complejidad de los datos que resulta en una situación de ajuste insuficiente. A medida que aumenta el índice de complejidad, el error alcanza su mínimo para el modelo óptimo h^* antes de que vuelva a aumentar. Para los índices de modelo alto, la estructura comienza a adaptar su modelo de aprendizaje a los datos de entrenamiento que resulta en un ajuste excesivo que reduce el valor de error de entrenamiento y aumenta el VC del modelo sin embargo, a expensas de un deterioro en el error de prueba, [Khanna and Awad, 2015].

3.4. Clasificador de margen máximo.

A continuación se definen conceptos que a menudo se usan indistintamente y se introduce el concepto de hiperplano de separación óptimo.

3.4.1. Hiperplano de separación

En un espacio $n - dimensional$, un hiperplano es un subespacio afín de dimensión $n - 1$. La definición matemática de un hiperplano es bastante simple y se expresa mediante la expresión:

$$b + w_1x_1 + w_2x_2 + ... + w_nx_n = 0 \tag{3.3}$$

Supongamos que $\mathbf{x} = (x_1, x_2, ..., x_n)$ no satisface la ecuación anterior, mas bien

$$b + w_1x_1 + w_2x_2 + ... + w_nx_n > 0,$$

entonces $\mathbf{x}$ está a un lado del hiperplano. Por otro lado, si

$$b + w_1x_1 + w_2x_2 + ... + w_nx_n < 0,$$

$\mathbf{x}$ está al otro lado del hiperplano. *Es decir, se puede determinar fácilmente en qué lado del hiperplano se encuentra un punto simplemente calculando el signo del lado izquierdo* de la ecuación 3.3.

3.4.2. Clasificación usando un hiperplano

Supongamos que tenemos las siguientes observaciones $(\mathbf{x}_i, y_i), i = 1, ..., l$, donde $\mathbf{x}_i \in R^n$ que caen en dos clases, esto es $y_i \in \{1, -1\}$. También tenemos una observación de prueba $\mathbf{x}^*$, el objetivo es hallar un clasificador basado en los datos de entrenamiento que clasifique correctamente la observación de prueba, para lograr esto se utilizará el enfoque que se basa en el concepto de un *hiperplano de separación*.

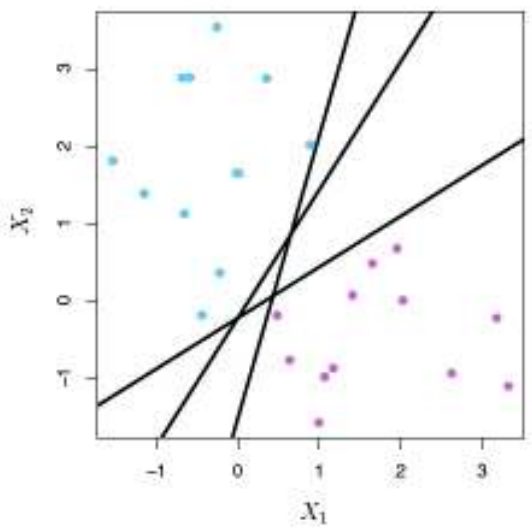

Figura 3.1: Observaciones que pertenecen a dos clases, que se muestran en azul y en morado. Además tres hiperplanos de separación de entre muchos posibles
Fuente: [Olkin, 2002]

Supongamos que es posible construir un hiperplano que separe las observaciones de entrenamiento perfectamente de acuerdo a sus etiquetas de clase. En la Figura 3.1 se indican tres *hiperplanos de separación*. Las observaciones azules se pueden etiquetar como $y_i = 1$ y las moradas como $y_i = -1$. El hiperplano de separación posee la propiedad para todo $\mathbf{x}_i$

$$\mathbf{w} \cdot \mathbf{x}_i + b > 0 \qquad si \qquad y_i = +1 \quad i = 1, ..., l$$

$$\mathbf{w} \cdot \mathbf{x}_i + b < 0 \qquad si \qquad y_i = -1 \quad i = 1, ..., l$$

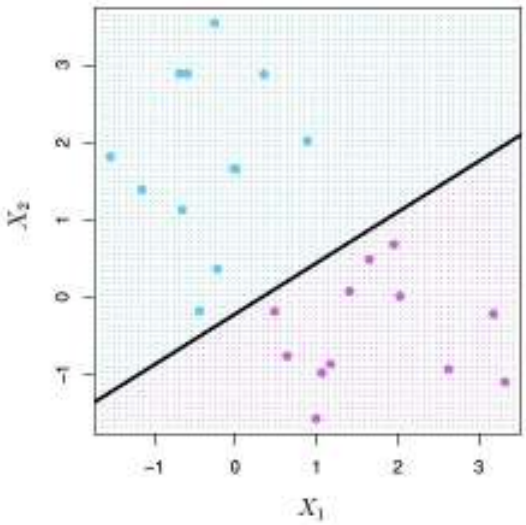

Figura 3.2: Hiperplano de separación
Fuente: [Olkin, 2002]

O también

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 0, \quad i = 1, ..., l$$

Si existe un hiperplano de separación, podemos usarlo para construir un clasificador muy natural: a una observación de prueba se le asigna una clase dependiendo de qué lado del hiperplano se encuentra. La Figura 3.2 muestra un ejemplo de dicho clasificador. Es decir, clasificamos la observación de prueba $\mathbf{x}^*$ basada en el signo de $D(\mathbf{x}^*) = (w_1 x_1^* + w_2 x_2^* + ... + w_n x_n^*) + b$. Si $D(\mathbf{x}^*)$ es positivo, entonces asignamos la observación de prueba a la clase 1, y si $D(\mathbf{x}^*)$ es negativo, entonces se la asigna a la clase -1. También podemos hacer uso de la magnitud de $D(\mathbf{x}^*)$. Si $D(\mathbf{x}^*)$ está lejos de cero, significa que $\mathbf{x}^*$ está lejos del hiperplano, por lo que podemos estar seguros sobre la asignación de clase de $\mathbf{x}^*$. Por otro lado, si $D(\mathbf{x}^*)$ está cerca de cero, entonces $\mathbf{x}^*$ está localizado cerca del hiperplano entonces estamos menos seguros acerca de la asignación de clase para $\mathbf{x}^*$.

3.4.3. Hiperplano de separación óptimo

En general, si los datos de entrenamiento se pueden separar perfectamente utilizando un hiperplano, entonces de hecho existirá un número infinito de tales hiperplanos. Para construir un clasificador basado en un hiperplano de separación, debemos tener una manera razonable de decidir cuál de los hiperplanos de separación usar.

Para lograr la separación máxima entre las dos clases, SVM elige el hiperplano que tiene el margen más grande. El margen es la suma de la distancia más corta desde el hiperplano de separación al punto de datos más cercano de ambas clases. Por tanto una opción natural es el *hiperplano de margen máximo* (también conocido como el *hiperplano de separación óptimo*), que es el hiperplano de separación más alejado de las observaciones de entrenamiento. Es decir, podemos calcular la

distancia (perpendicular) de cada observación de entrenamiento a un hiperplano de separación dado; la distancia más pequeña es la distancia mínima desde las observaciones al hiperplano, y se conoce como el *margen*. El hiperplano de margen máximo es el hiperplano de separación para el que el margen es más grande; es decir, es el hiperplano el que tiene la distancia mínima más alejada a las observaciones de entrenamiento. A continuación, podemos clasificar una observación de prueba en función de qué lado del hiperplano de margen máximo se encuentra.
Si $w_1, w_2, ..., w_n, b$ son los coeficientes del hiperplano de margen máximo, *entonces el clasificador de margen máximo clasifica la observación de prueba* $\mathbf{x}^*$ basado en el signo de $D(\mathbf{x}^*) = (w_1 x_1^* + w_2 x_2^* + ... + w_n x_n^*) + b$

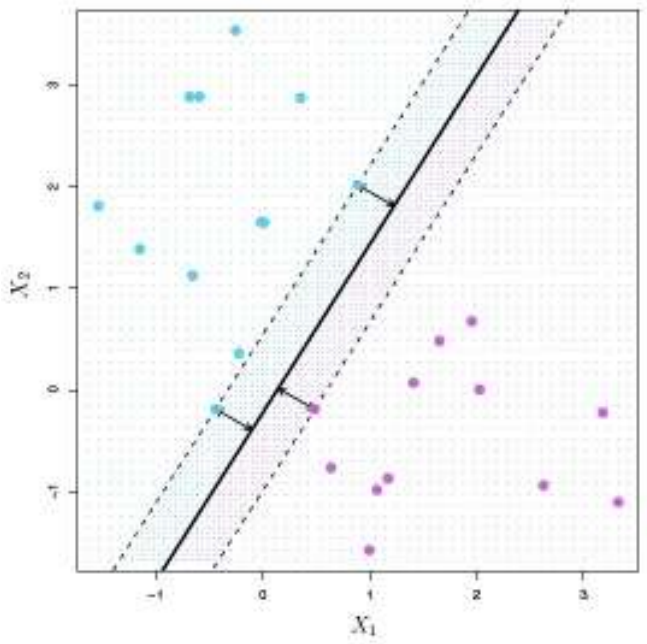

Figura 3.3: En este gráfico hay dos clases de observaciones, que se muestran en azul y en morado. El hiperplano de margen máximo se muestra como una línea continua. El margen es la distancia desde la línea continua a cualquiera de las líneas punteadas. Los dos puntos azules y el punto violeta que se encuentran en las líneas punteadas son los vectores de soporte, y la distancia desde esos puntos al hiperplano se indica con flechas.
Fuente: [Olkin, 2002]

La Figura 3.3 muestra el hiperplano de margen máximo en el conjunto de datos de la Figura 3.2. En cierto sentido, el hiperplano de margen máximo representa la línea media de la franja más ancha que podemos insertar entre las dos clases.
Al ver la Figura 3.3, vemos que tres observaciones de entrenamiento son equidistantes del hiperplano de margen máximo y se encuentran a lo largo de las líneas punteadas que indican el ancho del margen. Estas tres observaciones se conocen como *vectores de soporte,* ya que son vectores en el espacio n-dimensional y son el soporte del hiperplano de margen máximo en el sentido que, si estos puntos se mueven ligeramente, entonces el hiperplano de margen máximo se movería también.

El hiperplano de margen máximo depende directamente de los vectores de soporte, pero no de las otras observaciones: un movimiento a cualquiera de las otras observaciones no afectaría al hiperplano de separación, siempre que el movimiento de observación no lo haga cruzar el límite establecido por el margen.
El hiperplano de separación óptimo separa las dos clases y maximiza la distancia al punto más cercano de cualquiera de las clases. Esto no solo proporciona una solución única para el problema del hiperplano de separación, sino que al maximizar el margen entre las dos clases en los datos de entrenamiento, esto conduce a un mejor rendimiento de clasificación en los datos de prueba, [Shashua, 2009].

Una vez definidos estos tres conceptos, consideremos que según el tipo de separabilidad de los ejemplos de entrada, se consideran distintas opciones. Así, en primer lugar, se aborda el caso ideal de ejemplos perfectamente separables linealmente luego se abordará el caso más realista de ejemplos que, aunque afectados por ruido, se consideran linealmente cuasi-separables y, finalmente, se considera el caso de ejemplos no separables linealmente, donde las SVM demuestran su gran potencialidad. Es decir, se tienen los siguientes casos.

1. La clasificación binaria para ejemplos perfectamente separables mediante lo que se conoce como SVM de “margen rígido” (*hard margen*).

2. Dado que en la práctica es difícil tener el caso totalmente separable se tiene el problema de clasificación binaria para ejemplos cuasi-separables linealmente mediante SVM de “margen suave” (*soft margen*).

3. Para el caso de ejemplos no separables linealmente se utilizarán las funciones kernel.

[Fienberg and Casella,]

3.5. Máquina de soporte vectorial

3.5.1. Caso separable, SVM de margen rígido (SVM lineal)

El caso más simple de SVM son las máquinas lineales entrenadas en datos separables. El ***margen*** se define como la distancia del punto de entrenamiento más cercano al hiperplano de separación.
Sea un conjunto separable de ejemplos $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_l, y_l)\}$, donde $\mathbf{x}_i \in R^n$ y $y_i \in \{1, -1\}$, es decir, el conjunto de entrenamiento está compuesto de ejemplos positivos y negativos. La función de clasificación SVM $D(\mathbf{x})$ toma la forma $D(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} - b$, donde $\mathbf{w} \in R^n$ es el peso del vector $\mathbf{x}$ y es escalar b tal que:

$$\mathbf{w} \cdot \mathbf{x}_i - b \geq 1 \tag{3.4}$$

para ejemplos positivos y

$$\mathbf{w} \cdot \mathbf{x}_i - b \leq 1 \tag{3.5}$$

para ejemplos negativos.
Las condiciones anteriores pueden ser escritas como:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 \quad \forall(\mathbf{x}_i, y_i) \in S \tag{3.6}$$

Suponiendo que exista ese hiperplano, claramente no es único, en el caso separable hay más de un hiperplano de separación puesto que se puede desplazar o rotar sin entrar en contacto con las observaciones, la cuestión es encontrar una solución que tenga mejores propiedades de generalización, pues uno de los objetivos al construir una máquina de aprendizaje es obtener un buen rendimiento en el conjunto de prueba.
El signo de la distancia d_i de un punto $\mathbf{x}_i$ al hiperplano de separación está dado por:

$$d_i = \frac{\mathbf{w} \cdot \mathbf{x}_i + b}{\| \mathbf{w} \|} \tag{3.7}$$

Combinando 3.6 y 3.7, para todo $x_i \in S$ tenemos que el signo de la distancia d_i de un punto $\mathbf{x}_i$ al hiperplano de separación $(\mathbf{w}, b)$ está dado por:

$$y_i d_i \geq \frac{1}{\| \mathbf{w} \|} \tag{3.8}$$

Por tanto, $\frac{1}{\|\mathbf{w}\|}$ es el límite inferior de la distancia entre los puntos $\mathbf{x}_i$ y el hiperplano de separación $(\mathbf{w}, b)$.

Surge la pregunta por qué no simplemente reescribir la desigualdad 3.6 como:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 0 \quad \forall(\mathbf{x}_i, y_i) \in S$$

El propósito del 1 en el lado derecho de la desigualdad 3.6 es establecer una correspondencia uno a uno entre los hiperplanos de separación y su representación paramétrica. Esto se hace a través de la noción de representación canónica de un hiperplano de separación.

Definición 1. *Dado un hiperplano de separación $(\mathbf{w}, b)$ para todo conjunto separable S, la representación canónica de el hiperplano de separación es obtenido reescalando el par $(\mathbf{w}, b)$ al par $(\mathbf{w}', b')$ de tal manera que la distancia de el punto más cercano sea igual a $1/\|\mathbf{w}'\|$*

A través de esta definición tenemos que:

$$\text{mín}_{\mathbf{x}_i \in S}\{y_i(\mathbf{w}' \cdot \mathbf{x}_i + b')\} = 1$$

Por consiguiente, para un hiperplano de separación en la representación canónica, el límite en la desigualdad 3.7 es estrecho.
En lo que sigue asumiremos que un hiperplano de separación está dado en representación canónica y se escribirá como $(\mathbf{w}, b)$ en lugar de $(\mathbf{w}', b')$.

A partir de 3.8 se llega a la conclusión de que aumentar el margen es equivalente a disminuir la norma de $\mathbf{w}$.
[Pontil and Verri, 1998]

La búsqueda del hiperplano óptimo para el caso separable puede ser formalizado como el problema de encontrar el valor de $\mathbf{w}$ y b que minimiza $\|\mathbf{w}\|$ sujeto a las restricciones dadas por 3.6, es decir

$$\begin{aligned} &\text{mín} \qquad Q(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 \\ &s.a. \quad y_i(\mathbf{w}\cdot\mathbf{x}_i - b) - 1 \geq 0 \qquad \forall(\mathbf{x}_i, yi) \in S. \end{aligned} \tag{3.9}$$

Este problema de optimización con restricciones corresponde a un problema de programación cuadrático y es abordable mediante la teoría de la optimización. Dicha teoría establece que un problema de optimización, denominado primal, tiene una forma dual si la función a optimizar y las restricciones son funciones estrictamente convexas. En estas circunstancias, resolver el problema dual permite obtener la solución del problema primal. Para resolver el problema de optimización con restricciones 3.7 se utiliza los multiplicadores de Lagrange. Así la función objetivo es:

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i[y_i(\mathbf{w}\cdot\mathbf{x}_i - b) - 1], \tag{3.10}$$

donde las variables auxiliares no negativas α son los multiplicadores de Lagrange. La solución al problema de optimización restringida está determinada por el punto de silla de la función de Lagrange $J(\mathbf{w}, b, \alpha)$, que debe minimizarse con respecto a $\mathbf{w}$ y b; también debe ser maximizado con respecto a α. Por lo tanto, diferenciando $J(\mathbf{w}, b, \alpha)$ con respecto a $\mathbf{w}$ y b y estableciendo los resultados iguales a cero, obtenemos las siguientes dos condiciones de optimalidad:

$$Condición \quad 1: \qquad \frac{\partial J(\mathbf{w}, b, \alpha)}{\partial J\mathbf{w}} = 0 \tag{3.11}$$

$$Condición \quad 2: \qquad \frac{\partial J(\mathbf{w}, b, \alpha)}{\partial b} = 0 \tag{3.12}$$

Después de la reorganización de los términos, la Condición 1 se tiene:

$$\mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i \tag{3.13}$$

y de la Condición 2 se tiene:

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{3.14}$$

Para plantear el problema dual desarrollamos la Ecuación (7).

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}\cdot\mathbf{w} - \sum_{i=1}^{l} \alpha_i y_i \mathbf{w}\cdot\mathbf{x}_i - b\sum_{i=1}^{l} \alpha_i y_i + \sum_{i=1}^{l} \alpha_i \tag{3.15}$$

El tercer término del lado derecho de 3.15 es cero por la condición de optimalidad 3.14. Además, de la ecuación 3.13 se tiene:

$$\mathbf{w}\cdot\mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \mathbf{w}\cdot\mathbf{x} = \sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \mathbf{x}_j \tag{3.16}$$

En consecuencia, al establecer la función objetivo $J(\mathbf{w}, b, \alpha) = Q(\alpha)$, podemos reformular la ecuación 3.15 como:

$$Q(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \mathbf{x}_j, \tag{3.17}$$

donde α_i son no negativos.
Se establece el problema dual:

$$Q_1(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \mathbf{x}_j \tag{3.18}$$

$$sujeto \quad a \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{3.19}$$

$$\alpha \geq 0 \tag{3.20}$$

Una vez determinados los multiplicadores de Lagrange óptimos, denotados por α_i^*, podemos calcular el vector de pesos óptimo w^* utilizando la ecuación 3.13 y así tenemos:

$$\mathbf{w}^* = \sum_{i=1}^{l} \alpha_i^* y_i \mathbf{x}_i \tag{3.21}$$

De acuerdo con la propiedad de las condiciones de Kuhn-Tucker de la teoría de la optimización, la solución del problema dual α_i^* debe satisfacer la siguiente condición.

$$\alpha_i^* \{yi(\mathbf{w}^* \cdot \mathbf{x}_i - b) - 1\} = 0 \qquad para \quad todo \quad i = 1, ..., l \tag{3.22}$$

y ya sea α_i^* o su restricción correspondiente $\{yi(\mathbf{w}^* \cdot \mathbf{x}_i - b) - 1\}$ debe ser distinto de cero. Esta condición implica que solo cuando x_i es un vector de soporte o $\{yi(\mathbf{w}^* \cdot \mathbf{x}_i - b) = 1\}$, su coeficiente correspondiente α_i^* será distinto de cero. En otras palabras, los x_i cuyos coeficientes correspondientes α_i^* son cero no afectarán al vector de peso óptimo w^* debido a la ecuación 3.21. Por lo tanto, el vector de peso óptimo w^* solo dependerá de los vectores de soporte cuyos coeficientes no sean negativos. Una vez que calculamos el α_i^* no negativo y sus vectores de soporte correspondientes, podemos calcular el sesgo b usando un vector de soporte positivo x_i de la siguiente ecuación:

$$b^* = 1 - \mathbf{w}^* \cdot \mathbf{x}_i \tag{3.23}$$

La ecuación de clasificación se expresa de la siguiente forma:

$$F(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} - b \tag{3.24}$$

3.5.2. Caso cuasi separables, SVM de margen suave o clasificador de vector de soporte

Sin embargo, en muchos casos no existe un hiperplano de separación, por lo que no existe un clasificador de margen máximo. Un ejemplo se muestra en la

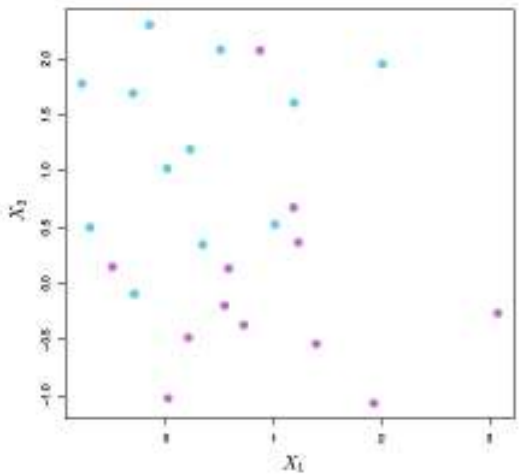

Figura 3.4: En este caso, las dos clases no son separables por un hiperplano, por tanto no se puede usar el clasificador de margen máximo.
Fuente: [Olkin, 2002]

Figura 3.4. En este caso, no podemos separar exactamente las dos clases.
Sin embargo, se puede extender el concepto de un hiperplano de separación para desarrollar un hiperplano que casi separe las clases, utilizando el llamado margen suave. La generalización del clasificador de margen máximo para el caso no separable se conoce como el *clasificador de vector de soporte*.

Para resolver el problema se puede relajar esta condición introduciendo el concepto de margen suave en el que la separabilidad se mantiene con un error.

$$\text{mín} \quad Q_1(\mathbf{w}, b, \xi_i) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \tag{3.25}$$

$$sujeto \quad a \quad y_i(\mathbf{w}\cdot\mathbf{x}_i - b) \geq 1 - \xi_i \qquad i = 1, ..., l \tag{3.26}$$

$$\xi_i \geq 0 \tag{3.27}$$

Debido a la presencia de ξ_i en 3.26 los datos de entrenamiento pueden clasificarse erróneamente, y la cantidad de errores de clasificación se minimizará al tiempo que se maximiza el margen de acuerdo con la función objetivo 3.25. C es un parámetro que determina la compensación entre el tamaño del margen y la cantidad de error en el entrenamiento.
De forma similar al caso de SVM de margen rígido, el problema primal se puede transformar a la forma dual utilizando los multiplicadores de Lagrange.

$$Q_2(\alpha) = \sum_{i=1}^{l}\alpha_i - \sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \mathbf{x}_i\mathbf{x}_j \tag{3.28}$$

$$sujeto \quad a: \quad \sum_{i=1}^{l}\alpha_i y_i = 0 \tag{3.29}$$

$$C \geq \alpha \geq 0 \tag{3.30}$$

El problema dual para el caso de patrones no separables es, por lo tanto, similar al del caso simple de patrones linealmente separables, excepto por una diferencia menor pero importante. La función objetivo $Q(\alpha)$ es la misma en los dos casos. El caso no separable difiere del caso separable en que la restricción $\alpha_i \geq 0$ se reemplaza con la restricción más estricta $C \geq \alpha_i \geq 0$. Excepto por esta modificación, la optimización restringida para el caso no separable y los cálculos de los valores óptimos del vector de ponderación $\mathbf{w}$ y sesgo b proceden de la misma manera que en el caso linealmente separable.
Del mismo modo que el SVM de margen rígido, α constituye una representación dual para el vector de ponderación de modo que,

$$\mathbf{w}^* = \sum_{i=1}^{l_s} \alpha_i^* y_i \mathbf{x}_i \tag{3.31}$$

donde l_s es el número de vectores de soporte cuyo coeficiente correspondiente $\alpha_i > 0$. La determinación de los valores óptimos del sesgo también sigue un procedimiento similar al descrito anteriormente. Una vez que α y b son calulados, la función en la ecuación 3.24 se usa para clasificar las nuevas observaciones.
Podemos decir aún más las relaciones entre α, ξ y C por las condiciones de Kuhn-Tucker que se definen por

$$\alpha_i\{y_i(\mathbf{w} \cdot \mathbf{x}_i) - 1 + \xi_i\} = 0, \qquad i = 1, ..., l \tag{3.32}$$

y

$$\mu_i \xi_i = 0, \qquad i = 1, ...l \tag{3.33}$$

En la ecuación 3.33, los multiplicadores de Lagrange μ_i se han introducido para hacer cumplir la no negatividad de las variables de holgura x_i para todos i. En el punto de silla de montar, la derivada de la función de Lagrange para el problema primario con respecto a la variable de holgura x_i es cero, de done se obtiene:

$$\alpha_i + \mu_i = C \tag{3.34}$$

De 3.33 y 3.34, se tiene que

$$\xi_i = 0 \quad si \quad \alpha_i < C, \quad y \tag{3.35}$$

$$\xi_i \geq 0 \quad si \quad \alpha_i = C \tag{3.36}$$

3.5.3. Caso no separable linealmente

A pesar que el margen máximo permite al SVM seleccionar entre varios hiperplanos candidatos, para muchos conjuntos de datos, el SVM puede no ser capaz de encontrar ningún hiperplano separador en absoluto porque los datos contienen instancias mal clasificadas. El problema se puede resolver utilizando un margen suave que acepte algunas clasificaciones erróneas de las instancias de entrenamiento.
Puesto que la mayoría de los problemas del mundo real implican datos no separables para los cuales no existe un hiperplano que separe con éxito las instancias positivas de negativas en el conjunto de entrenamiento. Una solución

al problema de inseparabilidad es mapear los datos en un espacio de mayor dimensión y definir allí un hiperplano de separación. Este espacio de mayor dimensión se denomina espacio de características, en oposición al espacio de entrada ocupado por las instancias de entrenamiento.
Con un espacio de características adecuadamente elegido de suficiente dimensionalidad, cualquier conjunto de entrenamiento consistente puede hacerse separable. Una separación lineal en el espacio de características corresponde a una separación no lineal en el espacio de entrada original. [Kotsiantis et al., 2006].

Sea $\mathbf{x}$ un vector en el espacio n-dimensional y $\varphi(\cdot)$ una función de mapeo no lineal desde el espacio de entrada al espacio de característica de alta dimensión. El hiperplano que representa el límite de decisión en el espacio de características se define de la siguiente manera.

$$\mathbf{w} \cdot \varphi(\mathbf{x}) - b = 0 \tag{3.37}$$

donde $\mathbf{w}$ denota el vector de ponderación que puede asignar los datos de entrenamiento en el espacio de características de alta dimensión al espacio de salida, y b es el sesgo. Usando la función $\varphi(\cdot)$, el peso se vuelve

$$\mathbf{w} = \sum \alpha_i y_i \varphi(\mathbf{x}_i). \tag{3.38}$$

La función de decisión llega a ser

$$F(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i y_i \varphi(\mathbf{x}_i) \cdot \mathbf{x} - b \tag{3.39}$$

Además, el problema dual de SVM de márgen suave ecuación 3.28 puede reescribirse utilizando la función de mapeo en los vectores de datos de la siguiente manera.

$$Q(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) \tag{3.40}$$

$$sujeto \quad a: \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{3.41}$$

$$C \geq \alpha \geq 0 \tag{3.42}$$

Tenga en cuenta que las funciones de asignación de características en el problema de optimización y también en la función de clasificación siempre aparecen como productos de puntos, por ejemplo, $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j \cdot \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j$ es el producto interno entre pares de vectores en el espacio de características transformadas. El cálculo del producto interno en el espacio de características transformadas parece ser bastante compleja y adolece del problema de la maldición de la dimensionalidad. Para evitar este problema, se usa el truco del kernel que reemplaza el producto interno en el espacio de características con una función kernel K en el espacio de entrada original de la siguiente manera.

$$K(\mathbf{u}, \mathbf{v}) = \varphi(\mathbf{u}) \varphi(\mathbf{v}) \tag{3.43}$$

El problema dual ahora se define utilizando la función kernel de la siguiente manera:

$$Q_3(\alpha) = \sum_{i=1}^{l} \alpha_i - \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{3.44}$$

$$sujeto \quad a: \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{3.45}$$

$$C \geq \alpha \geq 0 \tag{3.46}$$

La función de clasificación es:

$$F(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i), \mathbf{x}) - b \tag{3.47}$$

Las funciones kernel comúnmente utilizadas son:

- Lineal: $K(a, b) = a \cdot b$
- Polinomial: $K(a, b) = (a \cdot b + 1)^d$
- Función de base radial (RBF): $K(a, b) = exp(-\gamma \|a - b\|^2)$

[Yu and Kim, 2012]

3.6. El problema de desbalance en los datos

Los métodos de clasificación de los clientes suelen suponer que la distribución de clase de los datos de entrenamiento es equilibrada, es decir, que el número de muestras en cada clase es más o menos el mismo. Sin embargo, en problemas reales de clasificación de clientes, la distribución de clases de los datos de los clientes a menudo está desbalanceada.
La distribución de clase desbalanceada se caracteriza por tener muchas más instancias de algunas clases que otras. Particularmente para una aplicación de dos clases, el desequilibrio significa que una clase está representada por una gran cantidad de muestras, mientras que la otra está representada solo por unas pocas. En la calificación crediticia de los clientes, aquellos con poco crédito o consumidores fraudulentos suelen ser solo una pequeña fracción en todos los clientes. Los métodos de clasificación tradicionales obtienen un bajo rendimiento en los datos desbalanceados de clase, principalmente porque siempre se generalizan a partir del conjunto de capacitación en la hipótesis más simple que mejor se ajusta a los datos. Esta hipótesis rara vez se refiere a la muestra minoritaria de conjuntos de datos desequilibrados. Por lo tanto, las reglas de clasificación que predicen la clase minoritaria tienden a ser menores y más débiles que las de la clase mayoritaria. En consecuencia, las muestras de prueba pertenecientes a la clase minoritaria se clasifican erróneamente con mayor frecuencia que las pertenecientes a la clase mayoritaria. Sin embargo, en la mayoría de los campos de aplicación, el valor de la clasificación correcta para los clientes minoritarios es a menudo mayor que para los

de la mayoría. Por ejemplo, en la predicción de abandono de clientes, lo que más nos preocupa es si el modelo puede predecir correctamente los comportamientos de los clientes de abandono. Por lo tanto, cómo mejorar la capacidad de clasificación del algoritmo para los datos desequilibrados de clase es uno de los temas clave a resolver en la clasificación de los clientes [Xiao et al., 2012].

La deserción del cliente es a menudo un acontecimiento conocido como “raro.en las industrias de servicio, pero de gran interés y valor. Hasta hace poco el desbalance de clase no había recibido mucha atención en el contexto de la minería de datos. A continuación se presenta cómo manejar mejor el desequilibrio de clase en la predicción de deserción. La retención de clientes es una de las cuestiones más importantes para las empresas. La prevención de la deserción del cliente, como parte de un enfoque de Gestión de la Relación con el Cliente (CRM), está en las actividades de interés de las compañías. Las grandes compañías implementan modelos de predicción de deserción para poder detectar posibles desertores antes de que salgan efectivamente de la compañía. Sin embargo, hasta hace poco, el desequilibrio de clase no había recibido mucha atención en el contexto de la minería de datos. Ahora, a medida que se abordan los problemas del mundo real cada vez más complejos, el problema de los datos desequilibrados ocupa un lugar central.

De acuerdo a [Weiss, 2004] los problemas relacionados con datos no balanceados son:

1. *Métricas de evaluación inadecuadas:* a menudo, no se utilizan las mejores métricas para guiar los algoritmos de minería de datos y para evaluar sus resultados.

2. *Falta de datos, rareza absoluta:* el número de ejemplos asociados a la clase rara (minoritaria) es pequeño en un sentido absoluto, lo que dificulta la detección de regularidades dentro de la clase rara.

3. *Falta relativa de datos, rareza relativa:* los objetos no son raros en sentido absoluto, pero son raros en relación con otros objetos, lo que dificulta la heurística de búsqueda codiciosa, y los métodos globales en general no son manejables.

4. *Fragmentación de datos:* Muchos algoritmos de minería de datos, como los árboles de decisión, emplean un enfoque de división y conquista, donde el problema original se descompone en problemas cada vez más pequeños, lo que resulta en que el espacio de la instancia se divide en fragmentos más pequeños, lo cuál es un problema porque las regularidades sólo se pueden encontrar dentro de cada partición individual, la cuál contendrá menos datos.

5. *Sesgo inductivo inapropiado:* Generalizar a partir de ejemplos específicos, o inducción, requiere un sesgo extra evidencial. Sin ese sesgo, no son posibles “saltos inductivos” y el aprendizaje no puede ocurrir. El sesgo de un sistema de minería de datos es por lo tanto crítico para su rendimiento. Muchos aprendices utilizan un sesgo general con el fin de fomentar la generalización y evitar el sobre ajuste. Este sesgo puede afectar negativamente la capacidad de aprender casos raros y clases raras.

6. *Ruido:* Los datos ruidosos afectarán el funcionamiento de cualquier sistema de minería de datos, pero es interesante que el ruido tenga un mayor impacto en casos raros que en casos comunes.

3.6.1. Enfoques para afrontar el desbalance de datos

En la actualidad, hay dos enfoques comúnmente utilizados para abordar el problema del desequilibrio de clase en la clasificación de los clientes. Una es usar la técnica de remuestreo para equilibrar primero la distribución de clases y luego construir el modelo de clasificación en el conjunto de entrenamiento equilibrado. Los métodos de remuestreo más comúnmente utilizados incluyen submuestreo para muestras de la mayoría, muestreo excesivo para muestras minoritarias, método mejorado de muestreo SMOTE que es una combinación de submuestreo y más de sobremuestreo, [Xiao et al., 2012].
Según [Burez and Van den Poel, 2009], para afrontar el problema de desbalance de datos se ha propuesto el aprendizaje sensible al costo.

Aprendizaje sensible al costo

En muchas tareas de minería de datos, incluyendo la predicción de deserción, son los casos raros los que son de interés primario. Las métricas que no toman esto en cuenta generalmente no funcionan bien en estas situaciones. Una solución es utilizar métodos de aprendizaje sensibles al costo. Estos métodos pueden explotar el hecho de que el valor de identificar correctamente la clase positiva (rara) supera el valor de identificar correctamente la clase común. Para problemas de dos clases, esto se hace asociando un coste mayor con falsos negativos que con falsos positivos.
Asignar un mayor costo a falsos negativos que a falsos positivos mejorará el rendimiento con respecto a la clase de interés (rara). Si esta proporción de costos de clasificación errónea es 3: 1, entonces una región que tiene diez ejemplos negativos y cuatro ejemplos positivos, sin embargo, será etiquetada con la clase positiva. Por lo tanto, los costes no uniformes pueden influir en el clasificador para que tenga un buen rendimiento en la clase positiva, donde en este caso el sesgo es deseable.

Bosques aleatorios ponderados: Dado que el clasificador de los bosques al azar tiende a estar sesgado hacia la clase mayoritaria, se puede aplicar una penalización más severa al clasificar erróneamente a la clase minoritaria. Se asigna un peso a cada clase, con la clase minoritaria dada mayor peso (es decir, mayor costo de clasificación errónea). Los pesos de clase se incorporan en el algoritmo RF en dos lugares. En el procedimiento de inducción de árboles, los pesos de clase se utilizan para ponderar el criterio de Gini para encontrar splits. En los nodos terminales de cada árbol, los pesos de clase se toman de nuevo en consideración. La predicción de clase de cada nodo terminal se determina por “mayoría de votos ponderada”; es decir, el voto ponderado de una clase es el peso para esa clase multiplicado por el número de casos para esa clase en el nodo terminal. La predicción de clase final para RF se determina entonces agregando el voto ponderado de cada árbol individual, donde los pesos son pesos medios en los nodos terminales. Los pesos

de clase son un parámetro de ajuste esencial para lograr el rendimiento deseado. La estimación fuera de bolsa de la precisión de RF se puede utilizar para seleccionar pesos.

Muestreo

Una de las técnicas más comunes para tratar desbalance en los datos es el muestreo. La idea básica es eliminar o minimizar la rareza alterando la distribución de los ejemplos de entrenamiento. A continuación se explica un método de muestreo básico (submuestreo) y otro más avanzado (CUBE, un método de muestreo equilibrado eficiente).

- *Métodos básicos de muestreo:* Los métodos básicos de muestreo incluyen submuestreo y sobremuestreo. El submuestreo elimina ejemplos de clase mayoritaria, mientras que el sobre-muestreo, en su forma más simple, duplica ejemplos de la clase minoritaria. Ambas técnicas de muestreo disminuyen el nivel general de desequilibrio de clase.
  Debido a que el sobremuestreo introduce casos de entrenamiento adicionales, puede aumentar el tiempo necesario para construir un clasificador. Peor aún, debido a que el muestreo excesivo a menudo implica hacer copias exactas de los ejemplos, puede conducir a un ajuste excesivo. Además, el sobremuestreo no introduce nuevos datos, por lo que no aborda la cuestión fundamental de la falta de datos. Esto explica por qué algunos estudios han demostrado que el sobremuestreo simple es ineficaz para mejorar el reconocimiento de la clase minoritaria y por qué el submuestreo parece tener una ventaja sobre el muestreo excesivo.
  Por estas razones, en este trabajo se utilizará el submuestreo.

- *Métodos de muestreo avanzados:* Los métodos de muestreo avanzados pueden utilizar maneras para eliminar / añadir ejemplos o combinar las técnicas de sub-muestreo y de sobre-muestreo. El método CUBE es un método general que permite la selección de muestras aproximadamente equilibradas, ya que las estimaciones de Horvitz-Thompson para las variables auxiliares son iguales, o casi iguales, a los totales de su población. Este método es apropiado para un gran conjunto de variables de equilibrio cualitativo o cuantitativo, permite probabilidades de inclusión desiguales, y permite entender cuán precisamente una muestra puede ser equilibrada. Además, el diseño de muestreo respeta las probabilidades de inclusión fijas, iguales o desiguales. El método puede ser visto como una generalización del procedimiento de división que permite la construcción fácil de nuevos métodos de muestreo de probabilidad desigual.

Boosting

El objetivo principal del Boosting es mejorar el desempeño de las clasificaciones a través de la combinación de decisiones de muchos modelos de clasificación, llamados clasificadores débiles (o algoritmos de aprendizaje débiles). El boosting se ha aplicado con éxito a la predicción de deserción del cliente en las compañías

minoristas y de telecomunicaciones, [Vafeiadis et al., 2015].
Esta técnica sirve para mejorar la precisión de una función predictiva aplicando la función repetidamente en una serie y combinando la salida de cada función con la ponderación para que el error total de la predicción se minimice. En muchos casos, la precisión predictiva de tal serie excede en gran medida la exactitud de la función base utilizada. El primer algoritmo popular de este tipo fue AdaBoost, abreviado para el refuerzo adaptativo, por Freund y Schapire (1997). Este es un meta-algoritmo, y se puede utilizar junto con muchos otros algoritmos de aprendizaje para mejorar su rendimiento. AdaBoost es adaptativo en el sentido de que los clasificadores posteriores construidos se modifican a favor de aquellas instancias clasificadas erróneamente por clasificadores anteriores.

El *Gradient boosting* construye modelos de regresión aditiva ajustando secuencialmente una función parametrizada simple (base de aprendizaje) a los pseudo-residuales actuales por mínimos cuadrados en cada iteración. Los pseudo-residuos son el gradiente de la función de pérdida que se minimiza, con respecto a los valores del modelo en cada punto de datos de entrenamiento evaluado en la etapa actual. Se demuestra que tanto la precisión de aproximación como la velocidad de ejecución del aumento de gradiente pueden mejorarse sustancialmente incorporando la aleatorización al procedimiento. Específicamente, en cada iteración se toma una sub-muestra de los datos de entrenamiento al azar (sin reemplazo) del conjunto completo de datos de entrenamiento. Esta sub-muestra seleccionada al azar se utiliza entonces en lugar de la muestra completa para ajustarse al aprendizaje base y calcula la actualización del modelo para la iteración actual. Este enfoque aleatorio también aumenta la robustez frente a la sobrecapacidad del alumno de la base. Utilizando la conexión entre el impulso y la optimización, Friedman (2001) propone la máquina de elevación de gradiente.
En este trabajo se usó undersampling para afrontar el desbalance de datos.

3.6.2. Métricas de evaluación más apropiadas para el caso desbalanceado

Cuando se trata de un problema de clasificación binaria siempre se puede etiquetar una clase como positiva (en nuestro caso un desertor) y la otra como una clase negativa (no desertor). El conjunto de prueba consta de ejemplos positivos P y N negativos. Un clasificador asigna una clase a cada uno de ellos, pero algunas de las asignaciones son incorrectas. Para evaluar los resultados de la clasificación contamos el número de positivos verdaderos (TP), negativos verdaderos (TN), falsos positivos (FP) (en realidad negativos, pero clasificados como positivos) y falsos negativos (FN) (en realidad positivos, pero clasificados como negativos).

$$TP + FN = P, \tag{3.48}$$

y

$$TN + FP = N \tag{3.49}$$

En esta parte se explicará las métricas de evaluación más apropiadas para el caso de clasificación binaria que resultan a partir de una matriz de confusión Cuadro

3.1 [Burez and Van den Poel, 2009].
La definición de cada elemento de una matriz de confusión según [Kaur et al., 2013] es:

1. Verdaderos positivos (TP, True Positive): Número de casos positivos correctamente predichos.
2. Falsos Negativos (FN, False negative): Número de casos positivos pronosticados erróneamente como negativos.
3. Falsos positivos (FP, False positive): Número de casos negativos pronosticados erróneamente como positivos.
4. Verdadero negativo (TN, True negative):Número de casos negativos correctamente predichos.

Cuadro 3.1: Matriz de confusión para la evaluación de clasificadores

		Predicción de clase	
		Desertores	No desertores
Clase actual	Desertores	TP	FN
	No desertores	FP	TN

Con un conjunto de datos desbalanceado la métrica de precisión global en (3.53) ya no es suficiente. Para obtener una capacidad de clasificación equilibrada, generalmente se adoptan la *sensitividad* en (3.51) y la *especificidad* en (3.50) para monitorear el rendimiento de clasificación en dos clases por separado. La sensitividad también se llama tasa positiva verdadera o precisión positiva de clase, mientras que la especificidad se llama tasa negativa verdadera o precisión negativa de clase. Además, el área bajo curva ROC (AUC-ROC) indica una capacidad de clasificación equilibrada entre la sensibilidad y la especificidad en función de la variación de un umbral de clasificación.
Por otro lado, a veces estamos interesados en la capacidad de detección altamente efectiva para una sola clase. Por ejemplo, para un problema de detección de fraude con tarjeta de crédito, el objetivo es detectar transacciones fraudulentas. Para diagnosticar una enfermedad rara, lo que nos interesa especialmente es encontrar pacientes con esta enfermedad. Para tales problemas, a menudo se adopta otro par de métricas, precisión en (3.53) y sensitividad en (3.51). La medida-F en (3.57) se usa para integrar la precisión y la recuperación en una sola métrica para facilitar el modelado. De forma similar a AUC-ROC, el área bajo precisión / curva de recuperación (AUC-PR) se puede utilizar para indicar la capacidad de detección de un clasificador entre precisión y recuperación en función de la variación de un umbral de decisión, [Tang et al., 2009].

3.6.3. Definición de métricas para clasificación en el caso desbalanceado

A menudo es difícil o casi imposible construir un modelo de clasificación perfecto que clasifique correctamente todos los ejemplos del conjunto de pruebas. Por lo tanto, se tiene que elegir un modelo de clasificación subóptimo que mejor se adapte a las necesidades y funcione mejor en el dominio del problema. Se podría usar un clasificador que haga una predicción binaria (es decir, el cliente permanecerá con la empresa o no) o un clasificador que da una predicción de clase probabilística a qué clase pertenece un ejemplo. El primero se llama clasificador binario y el último se llama clasificador probabilístico. Además, se puede fácilmente convertir un clasificador probabilístico en uno binario usando un cierto umbral que separe las dos clases.

El clasificador asignó ejemplos $TP+FP$ a la clase positiva y $TN+FN$ ejemplos a la clase negativa. A continuación se define algunas medidas ampliamente utilizadas:

$$Especificidad = \frac{TN}{N} \Rightarrow 1 - especificidad = \frac{FP}{N} = tasa - FP, \tag{3.50}$$

$$Sensitividad = \frac{TP}{P} = tasa - TP = recall, \tag{3.51}$$

$$tasa - Y = \frac{TP + FP}{P + N}, \tag{3.52}$$

$$Precisión = \frac{TP}{TP + FP}, \tag{3.53}$$

$$Exactitud = \frac{TP + TN}{P + N}, \tag{3.54}$$

$$error \quad de \quad clasificación, MER = 1 - exactitud. \tag{3.55}$$

$$G - media = \sqrt{Sensitividad * Especificidad} \tag{3.56}$$

$$Medida - F = \frac{2 * Exactitud * Sensitividad}{Exactitud + Sensitividad} \tag{3.57}$$

Medidas como precisión, sesitividad y precisión (o MER) se utilizan a menudo para medir la calidad de clasificación de los clasificadores binarios. La tasa-FP o especificidad mide la fracción de no desertores que se clasifican erróneamente como desertores. La tasa-TP o sesitividad mide la fracción de desertores correctamente clasificada. Precisión mide la fracción de los casos clasificados como desertores que son verdaderamente desertores.
El lift indica cuánto mejor predice el clasificador en comparación con una selección aleatoria. Compara la precisión con la tasa de abandono general en el conjunto de prueba

$$Lift = \frac{precisión}{P/(P + N)} = \frac{Sensitividad}{TasaY} \tag{3.58}$$

ROC, ganancias acumuladas y curva de elevación: Cuando se desea evaluar la precisión de un clasificador independiente de cualquier umbral, se puede utilizar el análisis ROC. Un gráfico ROC se define mediante una definición paramétrica.

$$x = 1 - especificidad(t), \qquad y = sensitividad(t) \tag{3.59}$$

Cada clasificador binario (para un conjunto de ejemplos de prueba dado) está representado por un punto $(1 - especificidad, sensibilidad)$ en el gráfico. Variando el umbral del clasificador probabilístico, obtenemos un conjunto de clasificadores binarios, representados con un conjunto de puntos en el gráfico. La curva ROC es independiente de la relación P: N y por lo tanto es adecuada para comparar clasificadores cuando esta relación puede variar.
El área bajo la curva ROC se utiliza a menudo como una medida de la calidad de un clasificador probabilístico. Está cerca de la percepción de calidad de clasificación que la mayoría de la gente tiene. AUC se calcula con la siguiente fórmula:

$$AUC = \int_0^1 \frac{TP}{P} d\frac{FP}{N} = \frac{1}{PN} \int_0^N TPdFP \tag{3.60}$$

Un clasificador aleatorio tiene un área bajo la curva 0.5, mientras que un clasificador perfecto tiene 1. Los clasificadores utilizados en la práctica deberían por lo tanto estar en algún punto intermedio, preferiblemente cerca de 1, [Burez and Van den Poel, 2009].

Capítulo 4

MODELO DE DESERCIÓN

Esta sección se divide con base a los pasos necesarios para el aprendizaje y también haciendo uso de la metodología descrita en el 2.3.1.

4.1. Datos y supuestos

4.1.1. Comprensión del negocio

En la etapa del modelamiento es necesario conocer el contexto en que se está trabajando pues de esto depende el modelo de deserción adecuado que se implementará y además ayudará en la definición de la variable objetivo.

Para entender la importancia del contexto se considera los siguientes ejemplos y a continuación se detallará los posibles tipos de deserción que existen en el mercado de consumidores :

a. compras en un supermercado,
b. compras por catálogo,
c. servicios de guardería,
d. servicios de subscripción (revistas o tv cable).

En contextos tales como a y b, la empresa que provee el producto o servicio no conoce cuándo un cliente se convierte en desertor, puesto que tras realizar una compra, si un cliente no compra por un largo período de tiempo es difícil determinar si un cliente está inactivo o desertó. Una de las implicaciones de esta situación es que la empresa tiene que gestionar la retención de clientes y las compras para mejorar el valor de la vida útil de un cliente.
En contextos como c, servicios de guardería, la empresa sabría la certeza acerca de cuando un cliente deja de usar el servicio, pues sabe que todos los clientes desertarán en algún punto dentro de un tiempo sin importar los esfuerzos por retenerlos ya que los niños crecerán y dejará de usar sus servicios. Por tanto, en el caso c el objetivo de la empresa estará más centrado en estrategias de up-selling / cross-selling. De manera similar en contextos como d, servicios de subscripción, la empresa podría saber cuándo ocurre la deserción.
Tanto en c como en d, el conocimiento del tiempo de vida del cliente hace que sea

relativamente más fácil estimar el CLV.

El tipo de relación que existe entre una empresa y un cliente se puede dividir en **contractual** y **no contractual**. En el caso *contractual* los ingresos se pueden predecir con bastante precisión, en este caso las empresas conocen cuándo un cliente desertó. La relación cliente - proveedor es gobernada por un contrato o membresía. La duración de la relación cliente - empresa está estrechamente vinculada a la corriente de ingresos del cliente, por lo tanto, la vida útil del cliente está relacionada con el CLV de tal manera que una mayor vida útil del cliente significa mayor CLV para la empresa. Ejemplos de tales contextos serían servicios públicos, servicios de seguros, televisión por cable, servicios de suscripción de revistas y periódicos, algunos servicios de telefonía celular y algunas membresías de clubes.
Las relaciones *no contractuales* son aquellas en las que la empresa debe asegurarse que la relación se mantiene viva porque los clientes comúnmente dividen sus gastos de categoría entre varias empresas. En un contexto no contractual la empresa no conoce cuándo un cliente desertará, [Singh and Jain, 2010].

Por otro lado, los clientes desertores se dividen en dos grupos principales, desertores **voluntarios** y **no voluntarios**.

- La **deserción voluntaria** es la más difícil de determinar, porque este tipo de deserción ocurre cuando un cliente toma una decisión consciente de terminar su servicio con el proveedor. El abandono voluntario puede subdividirse en dos categorías principales, incidental y deliberado.

    - El abandono *incidental* ocurre cuando los cambios en las circunstancias impiden que el cliente requiera más el servicio proporcionado. Ejemplos de deserción incidental incluyen cambios en las circunstancias financieras del cliente, de modo que el cliente ya no puede pagar el servicio, o un traslado a una ubicación geográfica diferente donde el servicio de la compañía no está disponible. Por lo general, el abandono incidental sólo explica un pequeño porcentaje del abandono voluntario de una empresa.
    - El abandono *deliberado* es el problema que la mayoría de las soluciones de gestión de deserción tratan de atacar. Este tipo de deserción se produce cuando un cliente decide trasladar el uso de un servicio a una empresa competidora. Las razones que podrían conducir a una deserción deliberada del cliente incluyen razones basadas en la tecnología, cuando un cliente descubre que un competidor está ofreciendo los últimos productos, mientras que su proveedor actual no puede proporcionarlos. Las razones económicas incluyen encontrar el producto a un mejor precio de una compañía competidora. Ejemplos de otras razones para el rechazo deliberado incluyen factores de calidad tales como cobertura pobre, o posiblemente malas experiencias con centros de llamadas en el caso de telecomunicaciones, etc.

- Los **desertores no voluntarios** son más fáciles de identificar ya que estos son los clientes cuya relación termina por iniciativa de la empresa. Hay varias razones por las que una empresa podría revocar el servicio de un cliente, incluyendo el abuso de servicio y el no pago del servicio, [Hadden et al., 2007].

El contexto que se considerará para el modelamiento es el caso contractual, en el que los clientes tienen un acuerdo inicial, y después de un período deciden dejar de usar el servicio mediante un contrato o cierre de cuenta; y el tipo de deserción que se modelará es el caso voluntario deliberado.

4.1.2. Comprensión de los datos

Los datos que se usaron para el modelo de clasificación fueron proporcionados por una empresa del sector bancario.
Esta etapa podría ser la más relevante en la construcción de un modelo predictivo de abandono. La experiencia demuestra que la calidad y la idoneidad de los datos disponibles determinan la precisión y la potencia predictiva del modelo resultante. Diferentes combinaciones de datos pueden ser indicadores mejores o peores para diferentes problemas y para diferentes sectores. En última instancia, se trata de identificar los datos que mejor se ajusten al tipo de análisis que se está llevando a cabo, sólo así se podrá extraer, en etapas posteriores, un conocimiento útil y accesible en términos comerciales, [Garcia Gomez, 2014].

Hay varias formas de calcular la tasa de deserción, usualmente se lo calcula de la siguiente manera:

- Fijar un período convencional de tiempo como un mes o un año;
- Contar el número de clientes perdidos en este período;
- Dividir esta cantidad por el número de clientes que la empresa tenía al principio de este período.

En este trabajo la cantidad de clientes considerados en principio son 94411 de los cuáles 89112 son no desertores y 5299 desertores es decir un porcentaje desbalanceado de 94,39 % de no desertores frente a 5,61 % de desertores. Para afrontar el desbalance en los datos se uso submuestreo tras lo cual se obtuvo tres muestras balanceadas presentadas en el Cuadro 4.1, luego el conjunto de entrenamiento es el 80 % de la muestra y el conjunto de prueba es el 20 %.

4.1.3. Definición de deserción

El objetivo es *saber si un cliente de un banco cambiará o no de proveedor en un cierto período de tiempo.* Por lo tanto, el resultado sólo puede asumir dos valores para un cliente: desertó o no desertó. Esto sugiere que el problema puede ser afrontado como una tarea de clasificación, porque el dominio del resultado es finito.

Cuadro 4.1: Muestras obtenidas tras realizar submuestreo

		Conjunto de entrenamiento	*Conjunto de prueba*	*Totales*	*Porcentaje de desertores / no desertores en la muestra*
		11764	2941	14705	100,00 %
Muestra 1	**1**	4238	1053	5291	35,98 %
	0	7526	1888	9414	64,02 %
		11773	2943	14716	100,00 %
Muestra 2	**1**	4213	1073	5286	35,92 %
	0	7560	1870	9430	64,08 %
		11757	2939	14696	99,63 %
Muestra 3	**1**	4240	1046	5286	35,97 %
	0	7517	1839	9356	63,66 %

En particular, dado que el resultado tiene sólo dos valores posibles, el problema de predicción de deserción es una tarea de clasificación binaria.
La decisión acerca de cuánto tiempo de antemano queremos saber si un cliente desertará o no, depende de las necesidades de la empresa. En general, las pequeñas ventanas de tiempo permiten hacer predicciones más precisas, pero también reducen el tiempo de reacción del banco para tomar medidas de retención. Por otro lado, aunque las grandes ventanas de tiempo aumentan el tiempo permitido para la reacción del banco, podrían conducir fácilmente a resultados erróneos dado el alto tiempo de distancia. Todo esto sugiere que es necesario encontrar un buen consenso entre la exactitud de las predicciones y el tiempo de reacción permitido.

El concepto de tiempo es crítico en la predicción de la deserción. Este componente adicional se desvía de la tarea de clasificación típica, que se define como estática en sí. Por lo tanto, algunos trucos se utilizan a menudo en la literatura para representar el tiempo de una manera estática. La opción más simple consiste en representar a cada cliente con un registro, cuyas columnas son las características del cliente más importantes que afectan al evento de deserción. Estas características suelen seleccionarse con la ayuda del conocimiento del dominio y con árboles de decisión.

4.1.4. Propuesta de un marco de predicción de abandono dinámico

Generalmente, los clientes pueden ser caracterizados con diferentes tipos de atributos. Las características que no varían con el tiempo se llaman estáticas. Por ejemplo, las características estáticas de algunos clientes podrían ser: género, nivel de educación o estatus social. En particular, el nivel de educación varía muy lentamente o no varía en absoluto, por lo que se considera estático. Los atributos estáticos son fáciles de manejar.
Las características que dependen del tiempo se llaman atributos longitudinales o dinámicos. Describen los cambios de comportamiento de los clientes con el tiempo. Por lo tanto, son el ingrediente más poderoso para predecir la deserción del cliente. Los atributos dinámicos son más difíciles de gestionar que los estáticos, porque tienen que ser discretizados, para representarlos en un conjunto de datos. Esto

sugiere que debe elegirse una frecuencia de muestreo. Por ejemplo, si queremos modelar las visitas del cliente al banco, podemos contar sus visitas cada mes o cada semana. La frecuencia de muestreo es una decisión fundamental: si es demasiado baja, los datos no son suficientes para interpolar el comportamiento de los clientes. Por otro lado, si la frecuencia de muestreo es demasiado alta, muchos datos serán irrelevantes, porque los atributos longitudinales podrían variar con una tasa menor que la frecuencia de muestreo. Además, tener demasiados datos para analizar podría aumentar la complejidad. Normalmente, los atributos demográficos son considerados estáticos, mientras que los transaccionales son dinámicos, [Van den Poel and Lariviere, 2004].

En este trabajo la frecuencia de datos es mensual.

A continuación se presenta el marco de predicción de abandono dinámico que utiliza la naturaleza longitudinal de los datos del cliente de acuerdo [Ali and Arıtürk, 2014].

SPTD

El “marco estándar para la predicción de abandono del cliente” implica una observación por cliente con características que resumen la serie temporal histórica del comportamiento del cliente. Se define SPTD (datos de entrenamiento de período único) como un conjunto de datos de entrenamiento que consta de una observación por cliente que está activo a partir del tiempo de predicción, aunque las variables de predicción pueden resumir el historial del cliente.
En el caso de SPTD se tiene que es un conjunto de entrenamiento que contiene datos recopilados en un solo período de tiempo. Representa a cada cliente con un registro, cuyas características dinámicas se resumen en un solo valor con una función, como el valor medio o máximo, promedio o suma.

SPTD + retrasos

SPTD + retrasos se define como un conjunto de datos de entrenamiento que crea explícitamente k variables de retardo para cada variable de comportamiento del cliente además de las de SPTD. Sustancialmente, los SPTD + retrasos consideran k periodos de tiempo en los que se recopila información. Luego, los datos recopilados en los diferentes períodos de tiempo y se resumen con una función (como en el caso SPTD), creando k características para cada atributo dinámico.
Por ejemplo, puede contener el rendimiento anual de la cartera de inversiones del cliente calculado no solo en el último período, sino en los últimos k periodos.
La decisión sobre cuántos retrasos incluir es una variable de diseño similar a la cantidad de características de las series temporales de comportamiento del cliente para crear e incluir en el conjunto de datos de entrenamiento.

Preparación de los datos

Una vez que los datos son recolectados, tienen que ser organizados y transformados para definir un conjunto adecuado de características. Esto permite producir un conjunto de datos, que se utilizará para entrenar a los clasificadores. En esta sección mostramos cómo dividir y administrar los diferentes atributos de los clientes.

Como se mencionó anteriormente la frecuencia con la que se recopila los datos de los clientes es mensual. A continuación, la información es agrupada dentro de una única tabla , que tiene una clave primaria (es decir, cada registro está identificado de manera única por este campo) definido por el identificador del cliente y el mes considerado.
Según [Au et al., 2003] se sugiere una mejor manera de recopilar los atributos dinámicos relacionados con las transacciones. En particular, en lugar de considerar un período de tiempo fijo Figura 4.1 en el que se recopilan los datos, se sugiere utilizar un cronograma dinámico, como se muestra en las Figura 4.2. Por ejemplo: supongamos que un clasificador se construye utilizando datos de 1000 clientes, de los cuales 700 son no desertores y 300 son desertores. Supongamos también que se analizan las actividades de cinco meses, por ejemplo, de febrero de 2006 a junio de 2006. En este caso, el periodo de tiempo es fijo y solo se analizan las actividades realizadas en este intervalo. Supongamos que la mitad de los 300 clientes se han ido en abril. Esto implica que el modelo no estará completamente entrenado con sus datos, ya que sólo hay un mes de actividad disponible. Este problema se produce porque el período de tiempo es fijo. Por lo tanto, la solución sugerida es la siguiente:

1. decidir la duración del período de tiempo para analizar;

2. recopilar datos de la última fecha de transacción de cada cliente, con el fin de cubrir hacia atrás el período de tiempo establecido.

Aunque la predicción de la tasa de deserción del cliente es bien aceptada en la literatura, el enfoque típico no considera su variación en el tiempo, es decir, carece del aspecto dinámico. Por otra parte, la predicción dinámica de la deserción permitirá a la empresa asignar recursos de retención y mejora del servicio a través de clientes y tiempo. Desde el punto de vista de la gestión de la relación con el cliente, *los gerentes necesitan información sobre las razones y el momento de la interrupción para crear estrategias de retención y diagnosticar cuánto del desgaste puede ser controlado*. Además, la detección temprana de los desertores es valiosa, ya que proporciona tiempo a la empresa para convencer a esos clientes a permanecer antes de tomar su decisión final. Por último, las predicciones de deserción dinámicas son importantes para evaluar el valor agregado de la base de clientes de la empresa y, por lo tanto, su salud financiera.
El impacto del preprocesamiento de los datos de los clientes y la representación de los datos longitudinales de los clientes sobre la precisión predictiva es muy importante, por esta razón se propone un marco para la generación de datos de entrenamiento que facilite la predicción dinámica de la deserción, generando múltiples observaciones por cliente desde diferentes periodos de tiempo y, por lo tanto, incluyendo clientes que desertaron antes. Al hacer esta consideración en

los datos de entrenamiento se alivia el problema de eventos raros asociado con la predicción de deserción en muchas organizaciones [Burez and Van den Poel, 2009], lo que plantea dificultades para el aprendizaje de modelos de clasificación [Weiss, 2004]. [Ali and Arıtürk, 2014]. La norma establece que un cliente es “desertor” en

Figura 4.1: Línea de tiempo fija. Fuente [Bhambri, 2013]

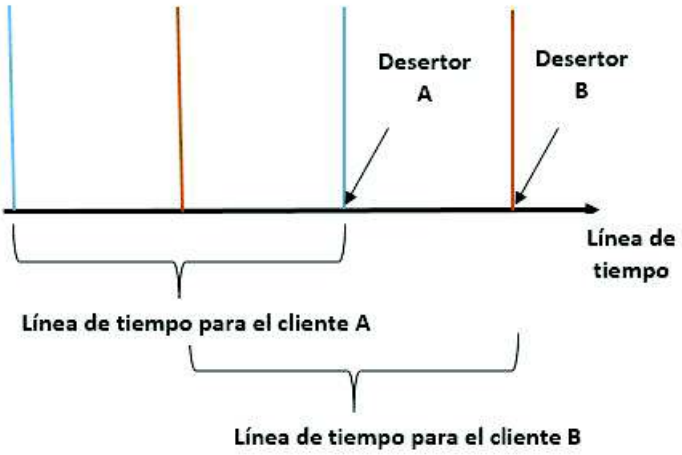

Figura 4.2: Línea de tiempo dinámica. Fuente [Bhambri, 2013]

un mes específico si cerró todas las cuentas bancarias en ese mes. Sin embargo, la necesidad principal de un banco es retener a sus clientes tanto tiempo como sea posible, es inútil identificar a los clientes con un aviso muy corto con respecto a su salida, ya que entonces el banco no puede rectificar el comportamiento de los clientes, período de tiempo no es suficientemente largo.

El objetivo es crear un clasificador capaz de detectar clientes que salen con suficiente anticipación. Sin embargo, la definición estándar de “desertor” no es suficiente para satisfacer este requisito, puesto que etiquetar los registros con esta definición crearía un clasificador útil para predecir el instante preciso de tiempo cuando el cliente se va, haciendo imposible aplicar acciones de rectificación

conductual de lealtad. Por lo tanto, es necesario modificar la definición estándar de “desertor”, para tener la seguridad que el clasificador predecirá un instante de tiempo lo suficientemente lejos del instante de deserción real.

En este trabajo se cambió el evento de deserción tres meses hacia atrás, *para etiquetar como desertores a aquellos clientes que salen durante los siguientes tres siguientes meses*. Esto significa que la etiqueta de deserción se activa para el antepenúltimo registro, mientras que los dos últimos registros se vuelven irrelevantes. De hecho, los datos de los desertores relacionados con los tres últimos meses introducirán ruido en el análisis, ya que estos datos no se refieren al evento que queremos predecir (es decir, el pronóstico de los clientes que saldrán durante los siguientes tres meses). Esto significa que es necesario eliminar a los clientes que se unen y luego abandonar el banco en menos de tres meses, porque no hay un período suficiente de tiempo disponible para aplicar las operaciones anteriores.
El uso de estas nuevas etiquetas para formar un clasificador conduce a la construcción de un modelo capaz de predecir si un cliente saldrá en los tres meses siguientes. De hecho, un clasificador buscará los valores de características que caracterizan el comportamiento de los clientes etiquetados como desertores con la nueva definición, es decir, los clientes que saldrán durante los siguientes tres meses.

4.1.5. Selección de variables

En situaciones del mundo real, las características relevantes a menudo se desconocen a priori. Por lo tanto, se presentan muchas características candidatas para representar mejor el dominio. Desafortunadamente, muchos de estos son parcial o completamente irrelevantes / redundantes para el objetivo.
En muchas aplicaciones, el tamaño de un conjunto de datos es tan grande que el aprendizaje puede no funcionar tan bien antes de eliminar estas características no deseadas. La reducción del número de características irrelevantes / redundantes reduce drásticamente el tiempo de ejecución de un algoritmo de aprendizaje y produce un concepto más general. Esto ayuda a obtener una mejor comprensión del concepto subyacente de un problema de clasificación del mundo real. Los métodos de selección de características intentan seleccionar un subconjunto de características que son relevantes para el concepto objetivo. La selección de características es definida por muchos autores mirándola desde varios ángulos. Pero como se esperaba, muchos de ellos son similares en intuición y / o contenido, [Dash and Liu, 1997].

En este trabajo se eligió CHAID (Chi-square Automatic Interaction Detector) Detector Automático de Interacción Chi-cuadrado, un método de inducción del árbol de decisión para clasificación. En cada paso, CHAID elige la variable independiente que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente. La variables utilizadas en SVM que resultaron significativas se indican en el Anexo A. Si el $p-valor$ es inferior a 0,05 se rechaza la hipótesis de independencia, es decir, que no hay relación entre las variables. Caso contrario la

relación entre la variable independiente y dependiente existe y es estadísticamente significativa.
El modelo CHAID es un método exploratorio de análisis de datos usado para estudiar las relaciones entre una variable dependiente y una serie grande de variables predictoras. El modelo desarrolla un árbol de clasificación con los “perfiles más importantes” formados por las variables independientes diferenciado los criterios de una variable dependiente. Los atributos que aparecen en las reglas de clasificación son indicadores objetivos que se encuentran en los datos y que se consideran influyentes en el concepto objetivo “deserción”. Por ejemplo, a partir de la regla de clasificación, el usuario puede concluir qué atributos son influyentes y relevantes para la clase objetivo de “deserción”, mientras que el resto de atributos no lo son, [Ng and Liu, 2000].

Como se indica en el nombre, CHAID utiliza pruebas de independencia Chi-cuadrado de Pearson, que evalúan una asociación entre dos variables categóricas. Un resultado estadísticamente significativo indica que las dos variables no son independientes, es decir, que existe una relación entre ellas. La fórmula del estadístico es:

$$\chi^2 = \sum_i \frac{(\text{observada}_i - \text{teorica}_i)^2}{\text{teorica}_i}$$

Cuanto mayor sea el valor de χ^2, menos verosímil es que la hipótesis nula (que asume la igualdad entre la distribución teórica y la observada) sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de chi-cuadrado, más ajustadas están ambas distribuciones.
Los grados de libertad gl vienen dados por :

$$gl = (r-1)(k-1),$$

donde r es el número de filas y k el de columnas.
El criterio de decisión es:

- No se rechaza H_0 cuando $\chi^2 < \chi^2_t(r-1)(k-1)$.
- En caso contrario se rechaza la hipótesis nula.

Donde t representa el valor proporcionado por las tablas, según el nivel de significación estadística elegido.

Las variables que se consideraron como significativas en las tres muestras se hallan en el Anexo A y aquellas que no se consideraron como explicativas son la situación laboral, la actividad económica a la que se dedica el cliente y si posee servicios de asistencia o no.

4.2. Representación

Es necesario considerar que en las aplicaciones de minería de datos incluso cuando el problema u objetivo es claro y enfocado, el proceso de minería sigue

siendo complicado, involucrando múltiples tareas a través de múltiples etapas. En nuestra aplicación, aunque estamos claros de que queremos retener a los clientes y el objetivo es “identificar los posibles desertores antes de que realmente deserten”, todavía es difícil saber por dónde empezar. Sin estudiar el dominio, es imposible para nosotros ir más lejos. El problema y la declaración de objetivos especificados a menudo no da ninguna pista en cuanto a qué tareas de minería de datos están involucrados, “cuáles” técnicas deben aplicarse, y “cómo” se aplican. Evidentemente, el problema de la retención del cliente debe descomponerse adicionalmente en varios sub problemas de tal manera que el conocimiento derivado de una tarea en una fase pueda servir como entrada para la siguiente fase. Dado que las técnicas y herramientas de minería de datos disponibles están diseñadas para ser específicas de las tareas y no específicas del problema, no pueden aplicarse directamente para resolver problemas del mundo real, [Ng and Liu, 2000].
En otras palabras esta etapa consiste en representar información relevante para la decisión de una forma que a los tomadores de decisión les quede clara las implicaciones. Esta parte se desarrolla mejor en la sección 4.6 con la aplicación al cálculo del valor del cliente.

4.3. Método y estimación

En esta sección se presenta el modelamiento, resultados y una aplicación del modelo de deserción.

4.3.1. Tiempo de predicción

Como hemos visto en la sección 4.1.4 si fijamos la línea de tiempo para todos los clientes, la calidad de los datos se reduce. En este trabajo se considera una línea de tiempo dinámica, que varía para cada cliente, se adapta una línea de tiempo dinámica dado que el tiempo de deserción varía de un cliente a otro de esta manera evitamos el problema de no entrenar adecuadamente el modelo.
Es decir, para clientes que desertaron se retrocedió 3 meses desde su fecha de deserción, a partir de ahí se fue calculando las sumas acumuladas trimestralmente, semestralmente, etc.
Para no desertores se fue calculando la sumas acumuladas hacia atrás desde la última fecha de transacción. El número de meses de datos a considerar para el análisis de deserción depende de las necesidades de cada empresa.
En este trabajo los datos son considerados desde enero 2015 hasta junio 2017, de los cuáles se descartan a los clientes cuyo ingreso fue menor a tres meses.

4.3.2. Modelamiento con regresión logística

El modelo de regresión normal está dado por:

$$P(X) = \alpha + \beta X, \quad donde: \quad X = (x_1, x_2, x_3, ..., x_n)$$

Para mantener la probabilidad $P(X)$ entre el límite 0 y 1, tenemos que usar la función logística, es decir:

$P(X) = \frac{e^{-(\alpha+\beta X)}}{1+e^{-(\alpha+\beta X)}}$

$Q(X) = 1 - P(X) = \frac{1}{1+e^{-(\alpha+\beta X)}}$

Los resultados de las probabilidades continuas que se producen mediante el modelo de regresión logística se discriminarán en dos grupos mediante el uso de un valor umbral. Por lo general, este valor umbral es 0,5, y en este trabajo se usó el valor determinado por el paquete R tal que controlan la tasa de falsos positivos y la tasa de falsos negativos maximizando el índice J de Youden, $J = Especificidad + sensibilidad - 1$.

Las variables que aportan a la explicación de la deserción del cliente y que se usaron para la creación de un modelo logístico se presentan en el Cuadro 4.1 las cuáles fueron elegidas usando el valor de divergencia KL (Kullback-Leibler), descrito en el Anexo F, hasta alcanzar un valor de $AUC = 0,9745$.

Cuadro 4.2: Variables consideradas en el modelo logístico. Elaborado por: Autor

SUM_ct_md_carteratotaltarjeta_T30_nuevo
SUM_ct_md_totalpasivo_T3_nuevo
SUM_ct_md_carteratotaltarjeta_T12_nuevo
SUM_ct_md_totalahorrosaldocontable_T6_nuevo
SUM_ct_md_totalproductoscliente_T6_nuevo
SUM_ct_md_totalahorrosaldopromedio_T6_nuevo
SUM_ct_md_totalahorrosaldopromedio_T9_nuevo

A partir del modelo de regresión se obtuvo las siguientes matrices de confusión con las tres muestra en los conjuntos de entrenamiento y prueba presentados en los Cuadros 4.3 y 4.4.

En el Cuadro 4.8 se presentan los resultados de exactitud, sensitividad y especifidad obtenidos con regresión logística y SVM obtenidos a partir de las matrices de confusión en las bases de prueba y entrenamiento.

4.3.3. Modelamiento con SVM

Para seleccionar el modelo con SVM se consideraron tres posibilidades en los núcleos: lineal, polinomial y RBF, a partir de los cuáles se elije el de mayor exactitud y considerando también sensibilidad y especificidad. Las matrices de confusión para cada kernel con los tres conjuntos de prueba y entrenamiento se presentan en los Cuadros 4.5 a 4.7.

Cuadro 4.3: Matrices de confusión obtenidas con regresión logística en tres conjuntos de prueba

Conjuntos de prueba

	Muestra 1		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 146
1	1742	119	Número de clientes desertores pronosticados como no desertores FN = 119
0	146	934	Número de clientes no desertores correctamente predichos TN = 934

Número de clientes desertores correctamente predichos TP = 1742

	Muestra 2		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 33
1	1837	148	Número de clientes desertores pronosticados como no desertores FN = 148
0	33	925	Número de clientes no desertores correctamente predichos TN = 925

Número de clientes desertores correctamente predichos TP = 1837

	Muestra 3		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 45
1	1848	119	Número de clientes desertores pronosticados como no desertores FN = 119
0	45	927	Número de clientes no desertores correctamente predichos TN = 927

Número de clientes desertores correctamente predichos TP = 1848

Cuadro 4.4: Matrices de confusión obtenidas con regresión logística en tres conjuntos de entrenamiento

Conjuntos de entrenamiento

	Muestra 1		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 311
1	7215	597	Número de clientes desertores pronosticados como no desertores FN = 597
0	311	3641	Número de clientes no desertores correctamente predichos TN = 3641

Número de clientes desertores correctamente predichos TP = 7215

	Muestra 2		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 165
1	7354	471	Número de clientes desertores pronosticados como no desertores FN = 471
0	165	3783	Número de clientes no desertores correctamente predichos TN = 3783

Número de clientes desertores correctamente predichos TP = 7354

	Muestra 3		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 180
1	7337	469	Número de clientes desertores pronosticados como no desertores FN = 469
0	180	3771	Número de clientes no desertores correctamente predichos TN = 3771

Número de clientes desertores correctamente predichos TP = 7337

4.4. Evaluación

Con el fin de evaluar el rendimiento de los clasificadores en la predicción de la deserción para diferentes esquemas con sus parámetros apropiados, utilizamos las medidas de *exactitud*, *sensitividad*, *especificidad*, calculadas a partir de los valores de la matriz de confusión que se muestran en los Cuadros 4.3 a 4.7. Los casos verdadero positivo y falso positivo se denotan como TP y FP, respectivamente, mientras que los casos verdaderos negativo y falso negativo se denotan como TN y FN.
La exactitud es la proporción del número total de predicciones que eran correctas y se calcula a partir de la ecuación (4.1).

$$Exactitud = \frac{TP + TN}{P + N}, \tag{4.1}$$

La sensitividad es la proporción de casos positivos que fueron correctamente identificados y se calcula a partir de la ecuación (4.2).

$$Sensitividad = \frac{TP}{P} \tag{4.2}$$

La especificidad mide la fracción de no desertores que se clasifican erróneamente como desertores, se calcula de acuerdo a 4.3.

$$Especificidad = \frac{TN}{N} \tag{4.3}$$

La sensitividad por sí solo no puede describir la eficiencia de un clasificador, ya que el buen desempeño en uno de esos índices no implica necesariamente un buen desempeño en el otro. Por esta razón, la medida-F, una combinación popular se utiliza comúnmente como una métrica única para evaluar el rendimiento del clasificador. La medida F se define como la media armónica de exactitud y sensitividad.

$$Medida - F = \frac{2 * Exactitud * Sensitividad}{Exactitud + Sensitividad} \tag{4.4}$$

Un valor más cercano a uno implica que el clasificador obtiene una mejor combinación de precisión y sensitividad. [Vafeiadis et al., 2015].

Las matrices de confusión obtenidas en este trabajo con los diferentes tipos de kernel se presentan en los Cuadros 4.5 a 4.10.

A partir de los Cuadros 4.3 a 4.10 se obtuvo la tabla de validación que se indica en el Cuadro 4.11.

4.5. Selección del modelo

Basándonos en el Cuadro 4.11 y 4.13 se concluye que el mejor modelo de predicción de deserción de clientes se logra usando máquinas de soporte vectorial

Cuadro 4.5: Matrices de confusión obtenidas con SVM kernel lineal en tres conjuntos de prueba

Conjunto de prueba

	Muestra 1		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 143
1	903	59	Número de clientes desertores pronosticados como no desertores FN = 59
0	143	1836	Número de clientes no desertores correctamente predichos TN = 1836

Número de clientes desertores correctamente predichos TP = 903

	Muestra 2		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 69
1	1004	37	Número de clientes desertores pronosticados como no desertores FN = 37
0	69	1833	Número de clientes no desertores correctamente predichos TN = 1833

Número de clientes desertores correctamente predichos TP = 1004

	Muestra 3		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 91
1	955	42	Número de clientes desertores pronosticados como no desertores FN = 42
0	91	1851	Número de clientes no desertores correctamente predichos TN = 1851

Número de clientes desertores correctamente predichos TP = 955

Cuadro 4.6: Matrices de confusión obtenidas con SVM kernel lineal en tres conjuntos de entrenamiento

Conjunto de entrenamiento

	Muestra 1		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 469
1	3776	191	Número de clientes desertores pronosticados como no desertores FN = 191
0	469	7328	Número de clientes no desertores correctamente predichos TN = 7328

Número de clientes desertores correctamente predichos TP = 3776

	Muestra 2		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 323
1	3890	127	Número de clientes desertores pronosticados como no desertores FN = 127
0	323	7433	Número de clientes no desertores correctamente predichos TN = 7433

Número de clientes desertores correctamente predichos TP = 3890

	Muestra 3		
	1	**0**	Número de clientes no desertores pronosticados como desertores FP = 299
1	3941	143	Número de clientes desertores pronosticados como no desertores FN = 143
0	299	7374	Número de clientes no desertores correctamente predichos TN = 7374

Número de clientes desertores correctamente predichos TP = 3941

Cuadro 4.7: Matrices de confusión obtenidas con SVM kernel radial en tres conjuntos de prueba

Conjunto de prueba

	Muestra 1	
	1	**0**
1	891	53
0	155	1842

Número de clientes desertores correctamente predichos TP = 891
Número de clientes no desertores pronosticados como desertores FP = 155
Número de clientes desertores pronosticados como no desertores FN = 53
Número de clientes no desertores correctamente predichos TN = 1842

	Muestra 2	
	1	**0**
1	975	28
0	98	1842

Número de clientes desertores correctamente predichos TP = 975
Número de clientes no desertores pronosticados como desertores FP = 98
Número de clientes desertores pronosticados como no desertores FN = 28
Número de clientes no desertores correctamente predichos TN = 1842

	Muestra 3	
	1	**0**
1	927	29
0	119	1864

Número de clientes desertores correctamente predichos TP = 927
Número de clientes no desertores pronosticados como desertores FP = 119
Número de clientes desertores pronosticados como no desertores FN = 29
Número de clientes no desertores correctamente predichos TN = 1864

Cuadro 4.8: Matrices de confusión obtenidas con SVM kernel radial en tres conjuntos de entrenamiento

Conjunto de entrenamiento

	Muestra 1	
	1	**0**
1	3684	186
0	561	7333

Número de clientes desertores correctamente predichos TP = 3684
Número de clientes no desertores pronosticados como desertores FP = 561
Número de clientes desertores pronosticados como no desertores FN = 186
Número de clientes no desertores correctamente predichos TN = 7333

	Muestra 2	
	1	**0**
1	4200	31
0	13	7529

Número de clientes desertores correctamente predichos TP = 4200
Número de clientes no desertores pronosticados como desertores FP = 13
Número de clientes desertores pronosticados como no desertores FN = 31
Número de clientes no desertores correctamente predichos TN = 7529

	Muestra 3	
	1	**0**
1	4227	28
0	13	7489

Número de clientes desertores correctamente predichos TP = 4227
Número de clientes no desertores pronosticados como desertores FP = 13
Número de clientes desertores pronosticados como no desertores FN = 28
Número de clientes no desertores correctamente predichos TN = 7489

Cuadro 4.9: Matrices de confusión obtenidas con SVM kernel polinomial en tres conjuntos de prueba

Conjunto de prueba

Muestra 1

	1	0
1	811	38
0	235	1857

Número de clientes desertores correctamente predichos TP = 811
Número de clientes no desertores pronosticados como desertores FP = 235
Número de clientes desertores pronosticados como no desertores FN = 38
Número de clientes no desertores correctamente predichos TN = 1857

Muestra 2

	1	0
1	918	22
0	155	1848

Número de clientes desertores correctamente predichos TP = 918
Número de clientes no desertores pronosticados como desertores FP = 155
Número de clientes desertores pronosticados como no desertores FN = 22
Número de clientes no desertores correctamente predichos TN = 1848

Muestra 3

	1	0
1	876	24
0	170	1869

Número de clientes desertores correctamente predichos TP = 876
Número de clientes no desertores pronosticados como desertores FP = 170
Número de clientes desertores pronosticados como no desertores FN = 24
Número de clientes no desertores correctamente predichos TN = 1869

Cuadro 4.10: Matrices de confusión SVM kernel polinomial en tres conjuntos de entrenamiento

Conjunto de entrenamiento

Muestra 1

	1	0
1	811	38
0	235	1857

Número de clientes desertores correctamente predichos TP = 811
Número de clientes no desertores pronosticados como desertores FP = 235
Número de clientes desertores pronosticados como no desertores FN = 38
Número de clientes no desertores correctamente predichos TN = 1857

Muestra 2

	1	0
1	3560	68
0	653	7492

Número de clientes desertores correctamente predichos TP = 3560
Número de clientes no desertores pronosticados como desertores FP = 653
Número de clientes desertores pronosticados como no desertores FN = 68
Número de clientes no desertores correctamente predichos TN = 7492

Muestra 3

	1	0
1	3584	90
0	656	7427

Número de clientes desertores correctamente predichos TP = 3584
Número de clientes no desertores pronosticados como desertores FP = 656
Número de clientes desertores pronosticados como no desertores FN = 90
Número de clientes no desertores correctamente predichos TN = 7427

Cuadro 4.11: Resultados del modelo en el conjunto de prueba

Conjunto de prueba					
Modelo	***Número de muestra***	***Sensitividad / Precisión***	***Especificidad***	***Exactitud***	***Medida - F***
Lineal	***Muestra1***	0,939	0,928	0,931	0,935
	Muestra2	0,964	0,964	0,964	0,964
	Muestra3	0,958	0,953	0,955	0,956
Polinomial	***Muestra1***	0,955	0,888	0,907	0,931
	Muestra2	0,977	0,923	0,940	0,958
	Muestra3	0,973	0,917	0,934	0,953
RBF	***Muestra1***	0,944	0,922	0,929	0,937
	Muestra2	0,972	0,949	0,957	0,965
	Muestra3	0,970	0,940	0,950	0,960
Lgístico	***Muestra1***	0,936	0,865	0,910	0,923
	Muestra2	0,925	0,966	0,938	0,932
	Muestra3	0,940	0,954	0,944	0,942

Cuadro 4.12: Resultados del modelo en el conjunto de entrenamiento

Conjunto de entrenamiento					
Modelo	***Número de muestra***	***Sensitividad / Precisión***	***Especificidad***	***Exactitud***	***Medida - F***
Lineal	***Muestra1***	0,952	0,940	0,944	0,948
	Muestra2	0,968	0,958	0,962	0,965
	Muestra3	0,965	0,961	0,962	0,964
Polinomial	***Muestra1***	0,955	0,888	0,907	0,931
	Muestra2	0,981	0,920	0,939	0,960
	Muestra3	0,976	0,919	0,937	0,956
RBF	***Muestra1***	0,952	0,929	0,937	0,944
	Muestra2	0,993	0,998	0,996	0,994
	Muestra3	0,993	0,998	0,997	0,995
Lgístico	***Muestra1***	0,924	0,921	0,923	0,923
	Muestra2	0,940	0,958	0,946	0,943
	Muestra3	0,940	0,954	0,945	0,942

con un kernel radial pues se obtiene mayor exactitud y una medida - F más cercana a 1.
Las curvas ROC (Receiver-Operating-Characteristic) del modelo SVM con kernel radial está dado por las Figura 4.3 - Figura 4.5 y el valor del área bajo la curva AUC (Area Under the Curve) se indican en el Cuadro 4.13.

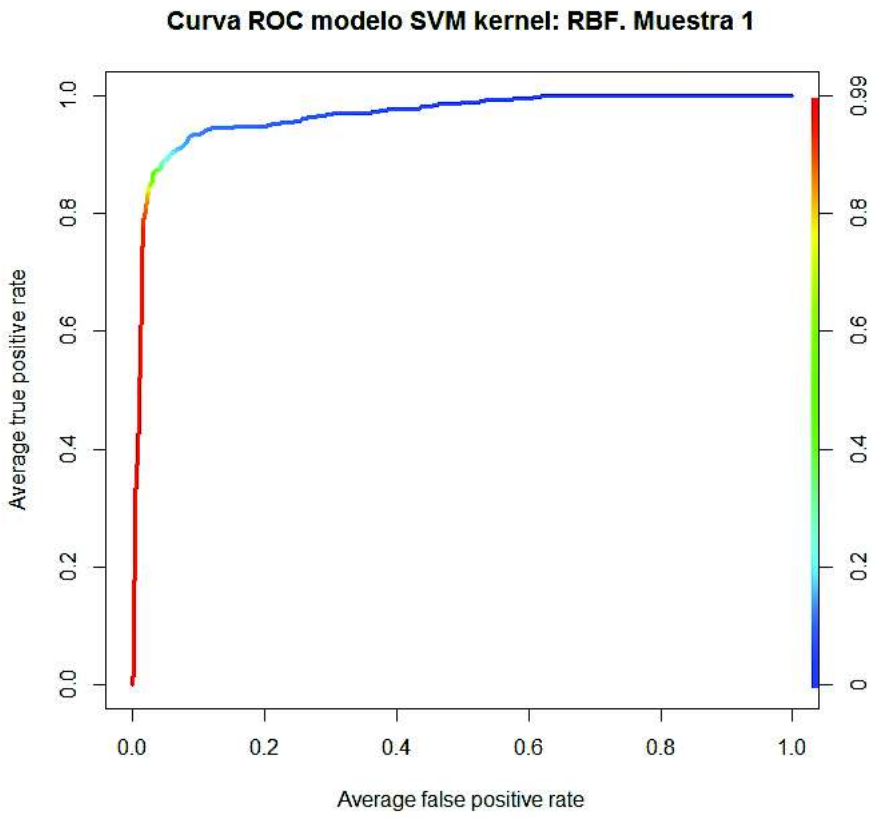

Figura 4.3: Curva ROC modelo SVM kernel: radial. Muestra 1. Elaborado por: Autor.

4.6. Aplicación al cálculo del CLV

La predicción de la deserción ayudará a una compañía a dirigir sus esfuerzos en la retención de clientes adecuados en el momento adecuado y así convencer a clientes valiosos a seguir usando servicios de la compañía. Además, la probabilidad de deserción es una parte importante en el cálculo del CLV [Ali and Arıtürk, 2014].

El sector bancario en particular debe operar con una estrategia de cliente a largo plazo, es decir, los clientes jóvenes son reconocidos como no rentables en la primera etapa en el ciclo de vida, pero se convertirán en rentables más adelante. Así como las relaciones de los clientes duran, tal vez décadas, la empresa debe abordar el valor de una pérdida potencial de un cliente. El análisis del valor de la vida útil del cliente ayudará a afrontar este desafío [Mutanen and Nousiainen, 2006].

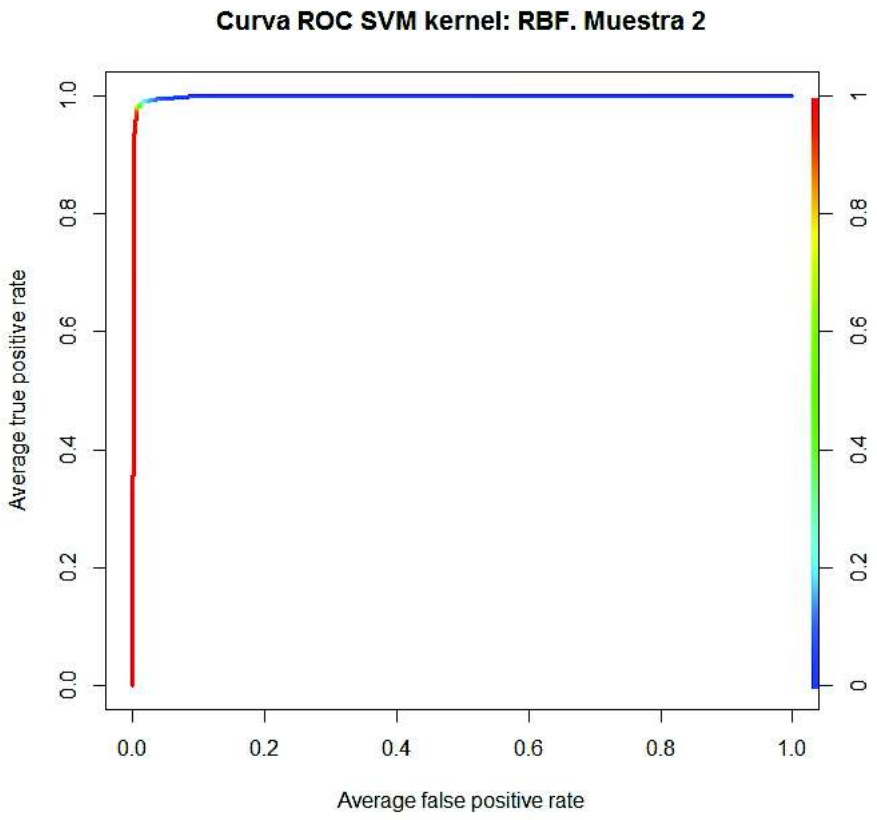

Figura 4.4: Curva ROC modelo SVM kernel: radial. Muestra 2. Elaborado por: Autor.

Los esfuerzos de gestión de la retención no deben centrarse en toda la base de clientes porque:

- no todos los clientes valen la pena retenerlos y,
- la retención de clientes que no tienen intención de desertar desperdiciará recursos, [Hadden et al., 2007]

En vista que el marketing es el *arte* de atraer y mantener clientes rentables, una empresa no debe tratar de perseguir y satisfacer a todo cliente. Por eso surge la siguiente pregunta: ¿qué hace a un cliente rentable?

Se definen un cliente rentable a una persona, familia o empresa cuyos ingresos con el tiempo superan en una cantidad aceptable los costos de la empresa de atraer, vender y dar servicio a ese cliente. Este exceso se llama valor de vida del cliente (CLV). EL cual debe ser una construcción importante en el diseño y el presupuesto de una serie de decisiones de marketing.

Para calcular el CLV se descuenta de los ingresos el costo de ventas como gastos de promoción incurridos para retener a los clientes tales como el envío de tarjetas de felicitación personalizadas, regalos y los gastos generales de promoción, con exclusión de las que se orientan directamente hacia la adquisición, se les

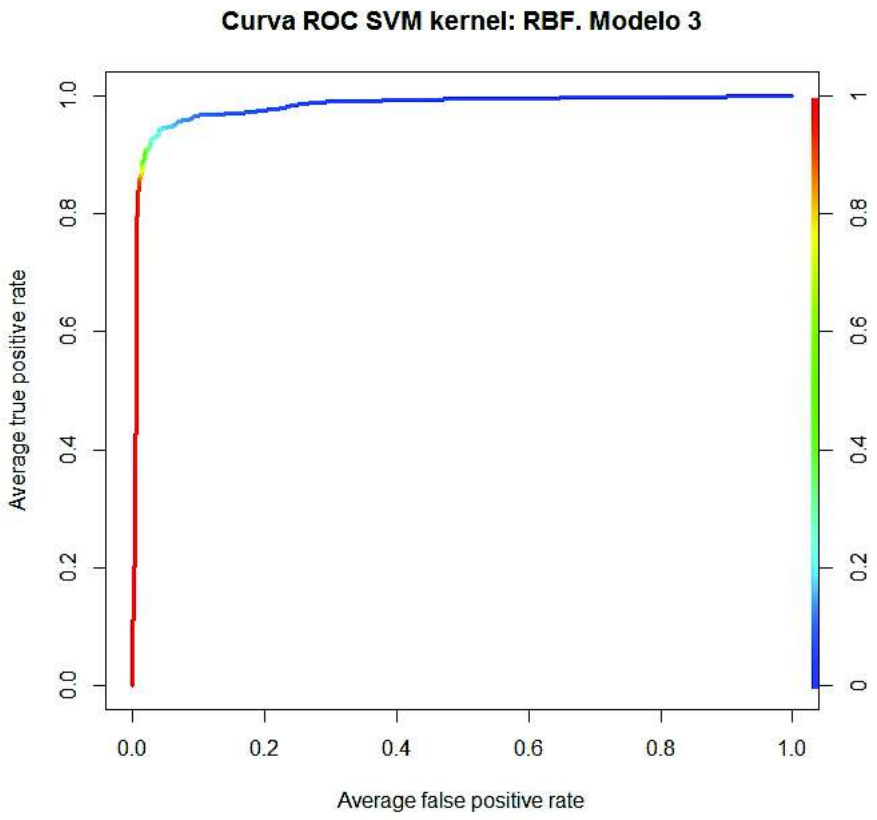

Figura 4.5: Curva ROC modelo SVM kernel: radial. Muestra 3. Elaborado por: Autor.

conoce como los costos de retención.

De acuerdo a [Mutanen and Nousiainen, 2006], la mayoría de los modelos CLV se derivan de la ecuación básica cuyos componentes son:

1. Tasa de retención o longitud de servicio
2. El valor actual neto de los clientes a lo largo del tiempo (ingresos y costos)
3. Factor de descuento.

Cada componente se puede medir o estimar por separado y luego combinarlo para el modelo CLV. El análisis del CLV también puede ayudar a la empresa en la personalización de productos y servicios. La comprensión del valor del cliente ayuda a la empresa a centrarse en los clientes productivos y en el segmento de clientes con potenciales impactos negativos a los ingresos.

El valor del CLV se estimó con la siguiente fórmula según [Blattberg and Neslin, 2008]:

$$CLV = (R - C)\frac{(1+\delta)}{(1+\delta-r)}, \tag{4.5}$$

donde:

Cuadro 4.13: Valores AUC obtenidos en los conjuntos de entrenamiento

Valores AUC obtenidos en los conjuntos de entrenamiento		
Regresión logística	***Muestra 1***	0,9745
	Muestra 2	0,9760
	Muestra 3	0,9689
SVM kernel lineal	***Muestra 1***	0,9602
	Muestra 2	0,9301
	Muestra 3	0,9754
SVM kernel RBF	***Muestra 1***	0,9659
	Muestra 2	0,9979
	Muestra 3	0,9821
SVM kernel polinomial	***Muestra 1***	0,9525
	Muestra 2	0,9776
	Muestra 3	0,9705

- r es la tasa de retención $r = 1 - probabilidad \quad de \quad desertar$,
- R y C son los ingresos y costos respectivamente que generará el cliente,
- δ es el factor de descuento. A pesar de la importancia del factor de descuento en el cálculo de LTV, hay poco trabajo sistemático sobre qué valor usar. En la práctica, los factores descuento anual que varían entre 10 % y 20 %.

Una vez que se obtuvo los cálculos del CLV la estrategia que se podría implementar para retener a clientes con alto valor podría ser comunicarse por teléfono y preguntarles su opinión relacionados con los siguientes puntos tomados de [Mihelis et al., 2001]:

1. *Personal del banco*: este criterio incluye todas las características relacionadas con el personal (habilidades y conocimiento, receptividad, comunicación y colaboración con los clientes, amabilidad, etc.).

2. *Imagen del banco:* la credibilidad del banco (nombre, reputación), la excelencia tecnológica, así como la capacidad de satisfacer las necesidades futuras de los clientes.

3. *Servicio:* este criterio se refiere al servicio ofrecido a los clientes; incluye la apariencia de los locales, el tiempo de espera (cola, teléfono, etc.), la complejidad de los procesos del servicio y la información provista (informando a los clientes de una manera comprensible, explicando el servicio y otros factores relevantes, informando sobre nuevos productos, etc.).

4. *Acceso:* la expansión de la red del banco, la ubicación de las sucursales, así como los problemas observados en el sistema de servicio como cajeros automáticos dañados.

En caso de haber recibido una queja enviar una tarjeta lamentando lo ocurrido. Además, se podría preguntarles acerca de sus preferencias en cuánto a compras y enviarles un bono con una cierta cantidad para usarlo en farmacias o supermercados.

4.7. Gestión de la rentabilidad tras una campaña de retención

Esta sección tiene como objetivo cuantificar la ganancia para la empresa que resulta de usar resultados obtenidos de deserción y CLV tras haber realizado una campaña de retención enfocada a clientes rentables.

Según [Neslin et al., 2006] a continuación se formula la rentabilidad de una campaña de gestión de deserción como una función de la capacidad del modelo predictivo para identificar posibles deserciones, para lo cual requiere de los siguientes datos:

- N : número total de clientes,
- α : fracción de clientes quienes se incluyen en el programa de retención,
- γ : fracción de clientes posiblemente desertores,
- δ : costo del incentivo del cliente para la empresa
- γ : fracción de desertores que deciden permanecer debido al incentivo (es decir, la tasa de éxito del incentivo),
- c : costo de contactar a un cliente para ofrecerle el incentivo,
- CLV : el valor de vida del cliente,
- A : costos administrativos fijos de ejecutar el programa de gestión de deserción.

Dadas estas definiciones, el beneficio que aporta una campaña de gestión de abandono es el siguiente:

$$\Pi = N\alpha[\beta\gamma(CLV - c - \delta) + \beta(1-\gamma)(-c) + (1-\beta)(-c-\delta)] - A \qquad (4.6)$$

El primer término entre corchetes refleja la contribución de ganancias entre la fracción $\beta\gamma$ de los clientes contactados que podrían ser desertores y deciden quedarse sobre la base del incentivo. El segundo término refleja el costo de contactar la fracción $\beta(1-\gamma)$ de posible desertores que no aceptan el incentivo y abandonan la empresa. El tercer término refleja el costo entre la fracción $(1-\beta)$ de clientes contactados que no son potenciales desertores pero aceptan la oferta. El término β refleja la exactitud del modelo:

- β_0 : la fracción de todos los clientes de la empresa que desertan.
- $\lambda(lift)$: Si $\lambda = 1$ significa que el modelo no proporciona esencialmente ningún poder de predicción. Como tal, este valor debería ser mayor que uno. Para calcular este valor se usó la ecuación 3.17.

Podemos expresar β como:

$$\beta = \lambda\beta_0 \tag{4.7}$$

Substituyendo 4.7 en 4.6 y reorganizando los términos, obtenemos:

$$\Pi = N\alpha[\gamma CLV + \delta(1-\gamma)]\beta_0\lambda - \delta - c - A \tag{4.8}$$

La ganancia incremental en el beneficio de un aumento de unidad en la precisión predictiva λ es la pendiente de la ecuación 4.7

$$Ganancia = N\alpha[\gamma CLV + \delta(1-\gamma)]\beta_0 \tag{4.9}$$

Capítulo 5

Conclusiones y Recomendaciones

5.1. Conclusiones

1. Como conclusión se tiene que es muy importante entender primeramente el contexto del negocio en el que se predecirá la deserción como la industria en la que se está trabajando con el fin de saber el tipo de deserción que se predecirá además de esto depende la definición de deserción.

2. Establecer un marco de predicción de abandono dinámico, en otras palabras establecer un tiempo de entrenamiento dinámico para cada cliente fue muy importante pues con este concepto nos aseguramos que el modelo esté bien entrenado. A diferencia de lo común que es establecer un tiempo de entrenamiento estático de 2 años para todos los clientes indistintamente.

3. La forma que se enfrentó el problema de desbalance de datos usando submuestreo proporcionó buenos resultados, además se concluye que las técnicas de aprendizaje automático produce mayor exactitud, sensitividad y especificidad frente a la regresión logística.

4. Se concluye que un modelo de deserción debe ir acompañado de un análisis del valor del cliente y de ganancia para la empresa pues de esa manera se aseguran que un cliente con alta probabilidad desertar tendrá un valor considerable para la empresa y por tanto concluir que vale la emprender una campaña de retención que genere beneficios.

5. Fue muy importante desechar de lo datos de entrenamiento a aquellos clientes cuyo tiempo de ingreso fue menor a tres meses ya que los mismos habrían introducido ruido en el modelo pues estos datos no se pertenecen al evento que queremos predecir.

6. Inicialmente considerando una muestra, el modelo SVM obtuvo mejores resultados con el kernel lineal. Sin embargo, al realizar el modelo con 2 muestras adicionales se concluyó que el kernel más conveniente para la predicción de la deserción es el RBF.

5.2. Recomendaciones

1. Se recomienda usar técnicas diferentes para afrontar el desbalance de datos descritos en la sección 3.6.1 como aprendizaje sensible al costo, sobremuestreo o técnicas de muestreo avanzado como CUBE con el fin de comparar la precisión obtenida con el submuestreo.

2. En este trabajo se realizó un marco de predicción de abandono dinámico, es decir la predicción de la deserción se hizo con tres meses de antelación para cada cliente dependiendo si es desertor o no. Se recomienda profundizar respecto a cuál sería el tiempo adecuado y necesario de antelación para la predicción de deserción.
   Además se recomienda comparar los resultados en el caso que la predicción se realice con un mes de antelación.

3. Introducir en el modelo información relacionada con RFM (Recency, Frecuency, Monetary), es decir frecuencia, tiempo reciente de uso de servicio y cantidad, como variables que ayuden a predecir la deserción.

4. Una vez obtenido el modelo de clasificación de deserción de clientes se recomienda usar técnicas de aprendizaje no supervisado descritos en la sección 2.2.5 para establecer patrones en los datos e identificar grupos de clientes con el fin establecer una estrategia de CRM eficaz y conducir a interacciones personalizadas con los clientes y, por lo tanto, una mayor satisfacción y relaciones con los clientes rentables a través del análisis de datos.

5. Se recomienda modelar los ingresos futuros de los clientes a quienes se emprenderá la campaña de retención para obtener el valor del CLV.

6. Para la selección de variables se recomienda usar otras técnicas de árboles de decisión como el método CART.

7. Profundizar en la selección del costo que se usará en el cada kernel al usar máquinas de soporte vectorial.

8. Calcular las ganancias que generan los clientes utilizando la 4.7 tras haber realizado una campaña de retención con valores aproximados a las necesidades de la empresa.

9. Se recomienda realizar la etapa de prueba en un conjunto de datos que pertenezcan a un tiempo diferente al de entrenamiento para calcular una exactitud más estricta del modelo.

10. Se recomienda realizar extracción de reglas que resultan del modelo SVM para fines de gestión de relaciones con clientes (CRM). Puesto que el principal inconveniente de SVM es que genera un modelo de “caja negra”, es decir, no revela el conocimiento aprendido durante el entrenamiento en forma comprensible, es necesario realizar el proceso de conversión de un modelo opaco en un modelo transparente a menudo se considera como extracción de reglas.

Anexos

Anexo A

Tablas de variables significativas mediante árboles de decisión

Tablas de variables explicativas elegidas mediante árboles de decisión.

Cuadro A.1: Variables significativas cuadro 1. Elaborado por: Autor

1	ct_cod_tipocliente_interno	**24**	SUM_ct_md_totalproductoscliente_T21
2	ct_cod_estadocliente	**25**	SUM_ct_md_totalproductoscliente_T24
3	ct_cod_genero	**26**	SUM_ct_md_totalproductoscliente_T27
4	ct_des_rango_edad	**27**	SUM_ct_md_totalproductoscliente_T30
5	ct_des_rango_antiguedad	**28**	SUM_ct_md_totalactivo_T3
6	ct_des_rango_ingreso	**29**	SUM_ct_md_totalactivo_T6
7	ct_cod_estado_civil	**30**	SUM_ct_md_totalactivo_T9
8	ct_des_nivel_academico	**31**	SUM_ct_md_totalactivo_T12
9	ct_cod_marca_visa	**32**	SUM_ct_md_totalactivo_T15
10	ct_cod_marca_crediflash	**33**	SUM_ct_md_totalactivo_T18
11	ct_cod_marca_cta_ahorro	**34**	SUM_ct_md_totalactivo_T21
12	ct_cod_marca_cta_corriente	**35**	SUM_ct_md_totalactivo_T24
13	ct_cod_marca_depplazo	**36**	SUM_ct_md_totalactivo_T27
14	ct_cod_marca_oportuna	**37**	SUM_ct_md_totalactivo_T30
15	ct_cod_marca_bgrnet	**38**	SUM_ct_md_totalpasivo_T3
16	ct_cod_marca_trnasfernet	**39**	SUM_ct_md_totalpasivo_T6
17	ct_cod_marca_otros	**40**	SUM_ct_md_totalpasivo_T9
18	SUM_ct_md_totalproductoscliente_T3	**41**	SUM_ct_md_totalpasivo_T12
19	SUM_ct_md_totalproductoscliente_T6	**42**	SUM_ct_md_totalpasivo_T15
20	SUM_ct_md_totalproductoscliente_T9	**43**	SUM_ct_md_totalpasivo_T18
21	SUM_ct_md_totalproductoscliente_T12	**44**	SUM_ct_md_totalpasivo_T21
22	SUM_ct_md_totalproductoscliente_T15	**45**	SUM_ct_md_totalpasivo_T24
23	SUM_ct_md_totalproductoscliente_T18	**46**	SUM_ct_md_totalpasivo_T27

Cuadro A.2: Variables significativas cuadro 2

47	SUM_ct_md_totalpasivo_T30	**70**	SUM_ct_md_carteraenriesgocredito_T9
48	SUM_ct_md_totalahorrosaldopromedio_T3	**71**	SUM_ct_md_carteraenriesgocredito_T12
49	SUM_ct_md_totalahorrosaldopromedio_T6	**72**	SUM_ct_md_carteraenriesgocredito_T15
50	SUM_ct_md_totalahorrosaldopromedio_T9	**73**	SUM_ct_md_carteraenriesgocredito_T18
51	SUM_ct_md_totalahorrosaldopromedio_T12	**74**	SUM_ct_md_carteraenriesgocredito_T21
52	SUM_ct_md_totalahorrosaldopromedio_T15	**75**	SUM_ct_md_carteraenriesgocredito_T24
53	SUM_ct_md_totalahorrosaldopromedio_T18	**76**	SUM_ct_md_carteraenriesgocredito_T27
54	SUM_ct_md_totalahorrosaldopromedio_T21	**77**	SUM_ct_md_carteraenriesgocredito_T30
55	SUM_ct_md_totalahorrosaldopromedio_T24	**78**	SUM_ct_md_totalahorrosaldocontable_T3
56	SUM_ct_md_totalahorrosaldopromedio_T27	**79**	SUM_ct_md_totalahorrosaldocontable_T6
57	SUM_ct_md_totalahorrosaldopromedio_T30	**80**	SUM_ct_md_totalahorrosaldocontable_T9
58	SUM_ct_md_totalcorrientesaldopromedio_T3	**81**	SUM_ct_md_totalahorrosaldocontable_T12
59	SUM_ct_md_totalcorrientesaldopromedio_T6	**82**	SUM_ct_md_totalahorrosaldocontable_T15
60	SUM_ct_md_totalcorrientesaldopromedio_T9	**83**	SUM_ct_md_totalahorrosaldocontable_T18
61	SUM_ct_md_totalcorrientesaldopromedio_T12	**84**	SUM_ct_md_totalahorrosaldocontable_T21
62	SUM_ct_md_totalcorrientesaldopromedio_T15	**85**	SUM_ct_md_totalahorrosaldocontable_T24
63	SUM_ct_md_totalcorrientesaldopromedio_T18	**86**	SUM_ct_md_totalahorrosaldocontable_T27
64	SUM_ct_md_totalcorrientesaldopromedio_T21	**87**	SUM_ct_md_totalahorrosaldocontable_T30
65	SUM_ct_md_totalcorrientesaldopromedio_T24	**88**	SUM_ct_md_totalcorrientesaldocontable_T3
66	SUM_ct_md_totalcorrientesaldopromedio_T27	**89**	SUM_ct_md_totalcorrientesaldocontable_T6
67	SUM_ct_md_totalcorrientesaldopromedio_T30	**90**	SUM_ct_md_totalcorrientesaldocontable_T9
68	SUM_ct_md_carteraenriesgocredito_T3	**91**	SUM_ct_md_totalcorrientesaldocontable_T12
69	SUM_ct_md_carteraenriesgocredito_T6	**92**	SUM_ct_md_totalcorrientesaldocontable_T15

Cuadro A.3: Variables significativas cuadro 3

93	SUM_ct_md_totalcorrientesaldocontable_T18	**116**	SUM_ct_md_carteratotalcredito_T12
94	SUM_ct_md_totalcorrientesaldocontable_T21	**117**	SUM_ct_md_carteratotalcredito_T15
95	SUM_ct_md_totalcorrientesaldocontable_T24	**118**	SUM_ct_md_carteratotalcredito_T18
96	SUM_ct_md_totalcorrientesaldocontable_T27	**119**	SUM_ct_md_carteratotalcredito_T21
97	SUM_ct_md_totalcorrientesaldocontable_T30	**120**	SUM_ct_md_carteratotalcredito_T24
98	SUM_ct_md_totaldepplazosaldocontable_T3	**121**	SUM_ct_md_carteratotalcredito_T27
99	SUM_ct_md_totaldepplazosaldocontable_T6	**122**	SUM_ct_md_carteratotalcredito_T30
100	SUM_ct_md_totaldepplazosaldocontable_T9	**123**	SUM_ct_md_carteratotaltarjeta_T3
101	SUM_ct_md_totaldepplazosaldocontable_T12	**124**	SUM_ct_md_carteratotaltarjeta_T6
102	SUM_ct_md_totaldepplazosaldocontable_T15	**125**	SUM_ct_md_carteratotaltarjeta_T9
103	SUM_ct_md_totaldepplazosaldocontable_T18	**126**	SUM_ct_md_carteratotaltarjeta_T12
104	SUM_ct_md_totaldepplazosaldocontable_T21	**127**	SUM_ct_md_carteratotaltarjeta_T15
105	SUM_ct_md_totaldepplazosaldocontable_T24	**128**	SUM_ct_md_carteratotaltarjeta_T18
106	SUM_ct_md_totaldepplazosaldocontable_T27	**129**	SUM_ct_md_carteratotaltarjeta_T21
107	SUM_ct_md_totaldepplazosaldocontable_T30	**130**	SUM_ct_md_carteratotaltarjeta_T24
108	SUM_ct_md_carteraenriesgotarjeta_T12	**131**	SUM_ct_md_carteratotaltarjeta_T27
109	SUM_ct_md_carteraenriesgotarjeta_T15	**132**	SUM_ct_md_carteratotaltarjeta_T30
110	SUM_ct_md_carteraenriesgotarjeta_T18	**133**	SUM_ct_md_numeroperaciones_activo_T3
111	SUM_ct_md_carteraenriesgotarjeta_T21	**134**	SUM_ct_md_numeroperaciones_activo_T6
112	SUM_ct_md_carteraenriesgotarjeta_T24	**135**	SUM_ct_md_numeroperaciones_activo_T9
113	SUM_ct_md_carteratotalcredito_T3	**136**	SUM_ct_md_numeroperaciones_activo_T12
114	SUM_ct_md_carteratotalcredito_T6	**137**	SUM_ct_md_numeroperaciones_activo_T15
115	SUM_ct_md_carteratotalcredito_T9	**138**	SUM_ct_md_numeroperaciones_activo_T18

Cuadro A.4: Variables significativas cuadro 4

139	SUM_ct_md_numeroperaciones_activo_T21
140	SUM_ct_md_numeroperaciones_activo_T24
141	SUM_ct_md_numeroperaciones_activo_T27
142	SUM_ct_md_numeroperaciones_activo_T30
143	SUM_ct_cod_marca_productocartera_T3
144	SUM_ct_cod_marca_productocartera_T6
145	SUM_ct_cod_marca_productocartera_T9
146	SUM_ct_cod_marca_productocartera_T12
147	SUM_ct_cod_marca_productocartera_T15
148	SUM_ct_cod_marca_productocartera_T18
149	SUM_ct_cod_marca_productocartera_T21
150	SUM_ct_cod_marca_productocartera_T24
151	SUM_ct_cod_marca_productocartera_T27
152	SUM_ct_cod_marca_productocartera_T30
153	ct_cod_actividad_economica_nueva

Anexo B

Código regresión logística en R

Las siguientes librerías se usaron para la regresión logística como para el modelo con SVM

```
library(e1071)
library(dplyr)
library(caret)
library(lift)
library(gains)
library(pROC)
library(ggvis)
library(gmodels)
library(rocc)
library(ROCR)
library(ggplot2)
library(cvAUC)
library(AUC)
library(InformationValue)
library(prediction)
library(lattice)

MODELO_LOG <- glm(
  formula = cancelado_a ~ SUM_ct_md_carteratotaltarjeta_T30_nuevo+SUM_ct_
      md_totalpasivo_T3_nuevo+SUM_ct_md_carteratotaltarjeta_T12_nuevo+
    SUM_ct_md_totalahorrosaldocontable_T6_nuevo+SUM_ct_md_
        totalproductoscliente_T6_nuevo+SUM_ct_md_totalahorrosaldopromedio_
        T6_nuevo+
    SUM_ct_md_totalahorrosaldopromedio_T9_nuevo+SUM_ct_md_
        totalahorrosaldocontable_T9_nuevo+SUM_ct_md_numeroperaciones_
        activo_T9_nuevo,
  family = binomial(),
  data = VARIABLES_ENTRENAMIENTO_LOG
)
length(MODELO_LOG$coefficients) > MODELO_LOG$rank
summary(MODELO_LOG)

predicciones_log_test <- predict(MODELO_LOG, VARIABLES_TEST_LOG,type = '
    response')
auc_log_test <-auc(roc(predicciones_log_test,VARIABLES_TEST_LOG$cancelado
    _a))
optCutOff <- optimalCutoff(VARIABLES_TEST_LOG$cancelado_a, predicciones_
    log_test, optimiseFor= "Both")
```

```
matriz_confusion_log_VARIABLES_TEST_LOG<-confusionMatrix(VARIABLES_TEST_
    LOG$cancelado_a, predicciones_log_test, threshold = optCutOff)
##################################---TEST CON LA BASE DE ENTRENAMIENTO
    ---#####################################

predicciones_log_train <- predict(MODELO_LOG, VARIABLES_ENTRENAMIENTO_LOG
    ,type = 'response')
auc_log_train <-auc(roc(predicciones_log_train,VARIABLES_ENTRENAMIENTO_
    LOG$cancelado_a))
optCutOff_train <- optimalCutoff(VARIABLES_ENTRENAMIENTO_LOG$cancelado_a,
     predicciones_log_train, optimiseFor= "Both")
matriz_confusion_log_VARIABLES_ENTRENAMIENTO_LOG<-confusionMatrix(
    VARIABLES_ENTRENAMIENTO_LOG$cancelado_a, predicciones_log_train,
    threshold = optCutOff_train)
```

Anexo C

Código SVM en R

```
##############################################*******MODELO SVM LINEAL****
    *##########################################
parametros_lineal <- tune.svm(cancelado ~., data=VARIABLES_SVM_
    ENTRENAMIENTO2, kernel='linear',cost=c(1,3,5))
summary(parametros_lineal)
parametros_lineal$best.parameters
summary(parametros_lineal)
svm_lineal_bp <- svm(cancelado ~., data=VARIABLES_SVM_ENTRENAMIENTO2,
    probability = TRUE,kernel='linear', cost=1, scale=FALSE)
summary(svm_lineal_bp)
print(svm_lineal_bp)
prediccion_lineal_BASE_TEST2<- predict(svm_lineal_bp, BASE_TEST2,decision
    .values = TRUE, probability = TRUE)
write.table(prediccion_lineal_BASE_TEST2, 'Borradores/BORRAR.txt', sep =
    '\t', quote = F, row.names = F)
test_lineal_BASE_TEST2 <- predict(parametros_lineal$best.model, BASE_
    TEST2)
matriz_confusion_lineal_BASE_TEST2<-table(test_lineal_BASE_TEST2, BASE_
    TEST2$cancelado)
resumen_matriz_confusion_lineal_BASE_TEST2<-confusionMatrix(matriz_
    confusion_lineal_BASE_TEST2)
##########################################-RESUTADOS svm_lineal CON LA
    BASE DE ENTRENAMIENTO-###########################

prediccion_lineal_c1_test <- predict(svm_lineal_bp, VARIABLES_SVM_
    ENTRENAMIENTO2,decision.values = TRUE, probability = TRUE)
test_train <- predict(parametros_lineal$best.model, VARIABLES_SVM_
    ENTRENAMIENTO2)
matriz_confusionlineal_VARIABLES_SVM_ENTRENAMIENTO2<-table(test_train,
    VARIABLES_SVM_ENTRENAMIENTO2$cancelado)
resumen_matriz_confusion_lineal_VARIABLES_SVM_ENTRENAMIENTO2<-
    confusionMatrix(test_train, VARIABLES_SVM_ENTRENAMIENTO2$cancelado)
##########---Gráfico modelo lineal---##############
svmmodel.class<-predict(svm_lineal_bp,BASE_TEST2, type="class")
svmmodel.labels<-BASE_TEST2$cancelado
svmmodel.probs <- attr(prediccion_lineal_BASE_TEST2, "probabilities")
    [1:2941,2]
svmmodel.prediction <- prediction(svmmodel.probs,svmmodel.labels)
svmmodel.performance<-performance(svmmodel.prediction,"tpr","fpr")
svmmodel.auc<-performance(svmmodel.prediction,"auc")@y.values[[1]]
```

```
curva_roc_lineal<-plot(svmmodel.performance, avg = "threshold", colorize
    = T, lwd = 3, main = "Curva ROC modelo SVM kernel: lineal. Muestra 1")
############******MODELO RBF*****##################
parametros_rbf <- tune.svm(cancelado ~., data=VARIABLES_SVM_
    ENTRENAMIENTO2,
                              kernel='radial', gamma = c(0.00009,0.001,0.1),
                                  cost = c(1,3,5))

summary(parametros_rbf)
parametros_rbf$best.parameters
svm_rbf_bp <- svm(cancelado ~., data=VARIABLES_SVM_ENTRENAMIENTO2,
    probability = TRUE, kernel='radial',gamma=0.001 ,cost=5)
summary(svm_rbf_bp)
prediccion_rbf_BASE_TEST2 <- predict(svm_rbf_bp, BASE_TEST2,decision.
    values = TRUE, probability = TRUE)
test_rbf_BASE_TEST2 <- predict(parametros_rbf$best.model, BASE_TEST2)
matriz_confusion_rbf_BASE_TEST2<-table(test_rbf_BASE_TEST2,BASE_TEST2$
    cancelado)
resumen_matriz_confusion_rbf_BASE_TEST2<-  confusionMatrix(test_rbf, BASE
    _TEST2$cancelado)
predicciones_rbf_valores_BASE_TEST2 <- attr(prediccion_rbf_BASE_TEST2, "
    decision.values")[1:2941,]
prediccines_rbf_probabilidades_BASE_TEST2<- attr(prediccion_rbf_BASE_
    TEST2, "probabilities")[1:2941,2]

#################################-RESUTADOS svm_RBF CON LA BASE DE
    ENTRENAMIENTO#####################
prediccion_RBF_VARIABLES_SVM_ENTRENAMIENTO2 <- predict(svm_rbf_bp,
    VARIABLES_SVM_ENTRENAMIENTO2,decision.values = TRUE, probability =
    TRUE)
test_RBF_VARIABLES_SVM_ENTRENAMIENTO2 <- predict(parametros_rbf$best.
    model, VARIABLES_SVM_ENTRENAMIENTO2)
matriz_confusion_RBF_VARIABLES_SVM_ENTRENAMIENTO2<-table(test_RBF_
    VARIABLES_SVM_ENTRENAMIENTO2,VARIABLES_SVM_ENTRENAMIENTO2$cancelado)
resumen_matriz_confusion_rbf_VARIABLES_SVM_ENTRENAMIENTO2<-
    confusionMatrix(test_RBF_VARIABLES_SVM_ENTRENAMIENTO2, VARIABLES_SVM_
    ENTRENAMIENTO2$cancelado)

##########---Gráfico modelo RBF---##############
svmmodel.class_rbf_BASE_TEST2<-predict(svm_rbf_bp,BASE_TEST2, type="class
    ")
svmmodel.labels_rbf_BASE_TEST2<-BASE_TEST2$cancelado
svm_rbf_bp$predictions<-predict(svm_rbf_bp,type='response',BASE_TEST2)
prediccion_rbf_BASE_TEST2
svmmodel.prediction_rbf_BASE_TEST2 <- prediction(prediccines_rbf_
    probabilidades_BASE_TEST2,svmmodel.labels_rbf_BASE_TEST2)
svmmodel.performance_rbf_BASE_TEST2<-performance(svmmodel.prediction_rbf_
    BASE_TEST2,"tpr","fpr")
svmmodel.auc_rbf_BASE_TEST2<-performance(svmmodel.prediction_rbf_BASE_
    TEST2,"auc")@y.values[[1]]
curva_roc_rbf<-plot(svmmodel.performance_rbf_BASE_TEST2, avg = "threshold
    ", colorize = T, lwd = 3, main = "Curva ROC modelo SVM kernel: RBF.
    Muestra 1" )

############******MODELO POLINOMIAL*****###############
parametros_poly <- tune.svm(cancelado ~., data=VARIABLES_SVM_
    ENTRENAMIENTO2,
```

```
                                kernel='polynomial', degree = 3, cost = c
                                    (1,3,5))

parametros_poly <- tune.svm(cancelado ~., data=VARIABLES_SVM_
    ENTRENAMIENTO2,
                                kernel='polynomial', degree = 3, cost = c
                                    (1,3,5))

summary(parametros_poly)
parametros_poly$best.parameters

svm_poly_bp <- svm(cancelado ~., data=VARIABLES_SVM_ENTRENAMIENTO2,
    probability = TRUE, kernel='polynomial' ,cost=5)
summary(svm_poly_bp)
prediccion_poly_BASE_TEST2 <- predict(svm_poly_bp, BASE_TEST2,decision.
    values = TRUE, probability = TRUE)
test_poly_BASE_TEST2 <- predict(parametros_poly$best.model, BASE_TEST2)
matriz_confusion_poly_BASE_TEST2<-table(test_poly_BASE_TEST2,BASE_TEST2$
    cancelado)
resumen_matriz_confusion_poly_BASE_TEST2<-confusionMatrix(test_poly_BASE_
    TEST2, BASE_TEST2$cancelado)
predicciones_poly_valores_BASE_TEST2<- attr(prediccion_poly_BASE_TEST2, "
    decision.values")[1:2941,]
prediccines_poly_probabilidades_BASE_TEST2<- attr(prediccion_poly_BASE_
    TEST2, "probabilities")[1:2941,2]
##############################-RESUTADOS svm_polinomial CON LA BASE DE
    ENTRENAMIENTO####################

prediccion_poly_c1_train <- predict(svm_poly_bp, VARIABLES_SVM_
    ENTRENAMIENTO2,decision.values = TRUE, probability = TRUE)
test_train_poly <- predict(parametros_poly$best.model, VARIABLES_SVM_
    ENTRENAMIENTO2)
matriz_confusion_poly_VARIABLES_SVM_ENTRENAMIENTO2<-table(test_train_poly
    ,VARIABLES_SVM_ENTRENAMIENTO2$cancelado)
##########---Gráfico modelo polinomial---##############
svmmodel.class_poly_BASE_TEST2<-predict(svm_poly_bp,BASE_TEST2, type="
    class")
svmmodel.labels_poly_BASE_TEST2<-BASE_TEST2$cancelado
svm_poly_bp$predictions<-predict(svm_rbf_bp,type='response',BASE_TEST2)
svmmodel.prediction_poly_BASE_TEST2 <- prediction(prediccines_poly_
    probabilidades_BASE_TEST2,svmmodel.labels_poly_BASE_TEST2)
svmmodel.performance_poly_BASE_TEST2<-performance(svmmodel.prediction_
    poly_BASE_TEST2,"tpr","fpr")
svmmodel.auc_poly_BASE_TEST2<-performance(svmmodel.prediction_poly_BASE_
    TEST2,"auc")@y.values[[1]]
plot(svmmodel.performance_poly_BASE_TEST2, avg = "threshold", colorize =
    T, lwd = 3, main = "Curva ROC modelo SVM kernel: Polinomial. Muestra 1
    ")
```

Anexo D

Coeficientes del hiperplano de separación

Pesos del hiperplano de separación, tabla 10.

Cuadro D.1: Pesos del hiperplano de separación, tabla 1. Elaborado por: Autor.

	Variable	Peso
1	ct_cod_tipocliente_interno_nuevo4_1	4,85195551
2	ct_cod_tipocliente_interno_nuevo4_2	-50,0903185
3	ct_cod_tipocliente_interno_nuevo4_3	45,238363
4	ct_des_rango_edad_nuevo11_2	-13,7983687
5	ct_des_rango_edad_nuevo11_3	40,3918793
6	ct_des_rango_edad_nuevo11_4	-1,59351067
7	ct_des_rango_antiguedad_nuevo14_2	38,801038
8	ct_des_rango_antiguedad_nuevo14_3	5
9	ct_des_rango_antiguedad_nuevo14_4	-7,84679998
10	ct_des_rango_antiguedad_nuevo14_5	-5,954238
11	ct_des_rango_ingreso_nuevo16_2	37,6860763
12	ct_des_rango_ingreso_nuevo16_3	43,3200208
13	ct_des_rango_ingreso_nuevo16_4	-4,10647553
14	ct_des_rango_ingreso_nuevo16_5	-46,8996216
15	ct_cod_estado_civil_nuevo18_2	1,4346715
16	ct_cod_estado_civil_nuevo18_3	13,4379681
17	ct_cod_estado_civil_nuevo18_4	0,20876597
18	ct_cod_estado_civil_nuevo18_5	14,9185945
19	ct_des_nivel_academico_nuevo19_2	12,0584721
20	ct_des_nivel_academico_nuevo19_3	14,0010527
21	ct_des_nivel_academico_nuevo19_4	-4,28506552
22	ct_des_nivel_academico_nuevo19_5	-25
23	SUM_ct_md_totalproductoscliente_T3_nuevo38_1_2	-268,476068
24	SUM_ct_md_totalproductoscliente_T3_nuevo38_1_3	41,53433
25	SUM_ct_md_totalproductoscliente_T3_nuevo38_1_4	-92,9382993
26	SUM_ct_md_totalproductoscliente_T3_nuevo38_1_5	-186,567548
27	SUM_ct_md_totalproductoscliente_T3_nuevoperdido	-36,903632
28	SUM_ct_md_totalproductoscliente_T6_nuevo38_2_2	-122,186277
29	SUM_ct_md_totalproductoscliente_T6_nuevo38_2_3	-26,6612984
30	SUM_ct_md_totalproductoscliente_T6_nuevo38_2_4	25,0629055
31	SUM_ct_md_totalproductoscliente_T6_nuevo38_2_5	-160,084283
32	SUM_ct_md_totalproductoscliente_T6_nuevo38_2_6	-46,9775334
33	SUM_ct_md_totalproductoscliente_T6_nuevoperdido	0,61294191
34	SUM_ct_md_totalproductoscliente_T9_nuevo38_3_1	247,984069
35	SUM_ct_md_totalproductoscliente_T9_nuevo38_3_2	-115,94105

Cuadro D.2: Pesos del hiperplano de separación, tabla 2.

36	SUM_ct_md_totalproductoscliente_T9_nuevo38_3_4	-115,477348
37	SUM_ct_md_totalproductoscliente_T9_nuevo38_3_5	-82,1550348
38	SUM_ct_md_totalproductoscliente_T9_nuevo38_3_6	76,6223061
39	SUM_ct_md_totalproductoscliente_T9_nuevo38_3_7	-61,4047763
40	SUM_ct_md_totalproductoscliente_T9_nuevoperdido	-10,8640684
41	SUM_ct_md_totalproductoscliente_T12_nuevo38_4_2	-143,639178
42	SUM_ct_md_totalproductoscliente_T12_nuevo38_4_3	-52,8255364
43	SUM_ct_md_totalproductoscliente_T12_nuevo38_4_4	-9,32261024
44	SUM_ct_md_totalproductoscliente_T12_nuevo38_4_5	-19,9258433
45	SUM_ct_md_totalproductoscliente_T12_nuevo38_4_6	50,9103702
46	SUM_ct_md_totalproductoscliente_T12_nuevoperdido	1,81872946
47	SUM_ct_md_totalproductoscliente_T15_nuevo38_5_2	-43,5084159
48	SUM_ct_md_totalproductoscliente_T15_nuevo38_5_3	-16,7748899
49	SUM_ct_md_totalproductoscliente_T15_nuevo38_5_4	4,61547367
50	SUM_ct_md_totalproductoscliente_T15_nuevo38_5_5	-25,208277
51	SUM_ct_md_totalproductoscliente_T15_nuevoperdido	-4,45044379
52	SUM_ct_md_totalproductoscliente_T18_nuevo38_6_2	-10,2275741
53	SUM_ct_md_totalproductoscliente_T18_nuevo38_6_3	-15,5544858
54	SUM_ct_md_totalproductoscliente_T18_nuevo38_6_4	-47,7346622
55	SUM_ct_md_totalproductoscliente_T18_nuevo38_6_5	-15,5118102
56	SUM_ct_md_totalproductoscliente_T18_nuevoperdido	3,70197949
57	SUM_ct_md_totalproductoscliente_T21_nuevo38_7_2	18,201397
58	SUM_ct_md_totalproductoscliente_T21_nuevo38_7_3	-43,6141859
59	SUM_ct_md_totalproductoscliente_T21_nuevo38_7_4	-33,8257617
60	SUM_ct_md_totalproductoscliente_T21_nuevo38_7_5	5,47369322
61	SUM_ct_md_totalproductoscliente_T21_nuevoperdido	-11,2351427
62	SUM_ct_md_totalproductoscliente_T24_nuevo38_8_2	-8,39954104
63	SUM_ct_md_totalproductoscliente_T24_nuevo38_8_3	-7,20856701
64	SUM_ct_md_totalproductoscliente_T24_nuevo38_8_4	-45,4573462
65	SUM_ct_md_totalproductoscliente_T24_nuevo38_8_5	0,83621904
66	SUM_ct_md_totalproductoscliente_T24_nuevoperdido	-4,77076475
67	SUM_ct_md_totalproductoscliente_T27_nuevo38_9_2	-11,7128451
68	SUM_ct_md_totalproductoscliente_T27_nuevo38_9_3	-21,4025791
69	SUM_ct_md_totalproductoscliente_T27_nuevo38_9_4	8,95941186
70	SUM_ct_md_totalproductoscliente_T27_nuevo38_9_5	-34,9833231
71	SUM_ct_md_totalproductoscliente_T27_nuevoperdido	-0,86066457
72	SUM_ct_md_totalproductoscliente_T30_nuevo38_10_2	-94,2227308
73	SUM_ct_md_totalproductoscliente_T30_nuevo38_10_3	39,4540526
74	SUM_ct_md_totalproductoscliente_T30_nuevo38_10_4	8,95941186
75	SUM_ct_md_totalproductoscliente_T30_nuevo38_10_5	-52,8725411
76	SUM_ct_md_totalproductoscliente_T30_nuevo38_10_6	45,1427665
77	SUM_ct_md_totalproductoscliente_T30_nuevoperdido	-11,4609591
78	SUM_ct_md_totalactivo_T3_nuevo39_1_2	-8,71051914
79	SUM_ct_md_totalactivo_T3_nuevo39_1_3	-9,0896306
80	SUM_ct_md_totalactivo_T3_nuevo39_1_4	39,4285272
81	SUM_ct_md_totalactivo_T3_nuevoperdido	-36,903632
82	SUM_ct_md_totalactivo_T6_nuevo39_2_2	-6,95082415
83	SUM_ct_md_totalactivo_T6_nuevo39_2_3	22,3850374
84	SUM_ct_md_totalactivo_T6_nuevo39_2_4	28,328087

Cuadro D.3: Pesos del hiperplano de separación, tabla 3.

86	SUM_ct_md_totalactivo_T9_nuevo39_3_2	-33,9967255
87	SUM_ct_md_totalactivo_T9_nuevo39_3_3	15,3182857
88	SUM_ct_md_totalactivo_T9_nuevo39_3_4	33,2928224
89	SUM_ct_md_totalactivo_T9_nuevoperdido	-10,8640684
90	SUM_ct_md_totalactivo_T12_nuevo39_4_2	-3,07895879
91	SUM_ct_md_totalactivo_T12_nuevo39_4_3	2,56097164
92	SUM_ct_md_totalactivo_T12_nuevo39_4_4	42,0712911
93	SUM_ct_md_totalactivo_T12_nuevoperdido	1,81872946
94	SUM_ct_md_totalactivo_T15_nuevo39_5_2	-0,69377567
95	SUM_ct_md_totalactivo_T15_nuevo39_5_3	-20,7195667
96	SUM_ct_md_totalactivo_T15_nuevo39_5_4	-20,890739
97	SUM_ct_md_totalactivo_T15_nuevoperdido	-4,45044379
98	SUM_ct_md_totalactivo_T18_nuevo39_6_2	-19,1707725
99	SUM_ct_md_totalactivo_T18_nuevo39_6_3	-25,7358211
100	SUM_ct_md_totalactivo_T18_nuevo39_6_4	-16,1899594
101	SUM_ct_md_totalactivo_T18_nuevoperdido	3,70197949
102	SUM_ct_md_totalactivo_T21_nuevo39_7_2	-4,75045525
103	SUM_ct_md_totalactivo_T21_nuevo39_7_3	-14,569216
104	SUM_ct_md_totalactivo_T21_nuevo39_7_4	-2,10343461
105	SUM_ct_md_totalactivo_T21_nuevoperdido	-11,2351427
106	SUM_ct_md_totalactivo_T24_nuevo39_8_2	9,72916568
107	SUM_ct_md_totalactivo_T24_nuevo39_8_3	-13,4986217
108	SUM_ct_md_totalactivo_T24_nuevo39_8_4	-3,36399765
109	SUM_ct_md_totalactivo_T24_nuevoperdido	-4,77076475
110	SUM_ct_md_totalactivo_T27_nuevo39_9_2	14,7291657
111	SUM_ct_md_totalactivo_T27_nuevo39_9_3	-3,08097273
112	SUM_ct_md_totalactivo_T27_nuevo39_9_4	10,7785089
113	SUM_ct_md_totalactivo_T27_nuevoperdido	-0,86066457
114	SUM_ct_md_totalactivo_T30_nuevo39_10_2	-1,12330311
115	SUM_ct_md_totalactivo_T30_nuevo39_10_3	-5,28855564
116	SUM_ct_md_totalactivo_T30_nuevo39_10_4	-4,51382877
117	SUM_ct_md_totalactivo_T30_nuevoperdido	-11,4609591
118	SUM_ct_md_totalpasivo_T3_nuevo40_1_2	246,451966
119	SUM_ct_md_totalpasivo_T3_nuevo40_1_3	166,57979
120	SUM_ct_md_totalpasivo_T3_nuevo40_1_4	161,588829
121	SUM_ct_md_totalpasivo_T3_nuevo40_1_5	38,6669904
122	SUM_ct_md_totalpasivo_T3_nuevoperdido	-36,903632
123	SUM_ct_md_totalpasivo_T6_nuevo40_2_2	5,06870484
124	SUM_ct_md_totalpasivo_T6_nuevo40_2_3	16,1065467
125	SUM_ct_md_totalpasivo_T6_nuevo40_2_4	18,5414143
126	SUM_ct_md_totalpasivo_T6_nuevo40_2_5	52,0028757
127	SUM_ct_md_totalpasivo_T6_nuevoperdido	0,61294191
128	SUM_ct_md_totalpasivo_T9_nuevo40_3_1	75,5081177
129	SUM_ct_md_totalpasivo_T9_nuevo40_3_2	10,5038015
130	SUM_ct_md_totalpasivo_T9_nuevo40_3_3	-29,968046
131	SUM_ct_md_totalpasivo_T9_nuevoperdido	-10,8640684
132	SUM_ct_md_totalpasivo_T12_nuevo40_4_2	-4,26313132
133	SUM_ct_md_totalpasivo_T12_nuevo40_4_3	-33,9523894
134	SUM_ct_md_totalpasivo_T12_nuevo40_4_4	-2,45519266

Cuadro D.4: Pesos del hiperplano de separación, tabla 4.

136	SUM_ct_md_totalpasivo_T12_nuevo40_4_6	23,490903
137	SUM_ct_md_totalpasivo_T12_nuevoperdido	1,81872946
138	SUM_ct_md_totalpasivo_T15_nuevo40_5_1	104,820717
139	SUM_ct_md_totalpasivo_T15_nuevo40_5_2	-12,2193835
140	SUM_ct_md_totalpasivo_T15_nuevo40_5_3	2,86109741
141	SUM_ct_md_totalpasivo_T15_nuevoperdido	-4,45044379
142	SUM_ct_md_totalpasivo_T18_nuevo40_6_1	74,6784404
143	SUM_ct_md_totalpasivo_T18_nuevo40_6_2	41,4203541
144	SUM_ct_md_totalpasivo_T18_nuevo40_6_3	-39,2233259
145	SUM_ct_md_totalpasivo_T18_nuevoperdido	3,70197949
146	SUM_ct_md_totalpasivo_T21_nuevo40_7_1	73,4341208
147	SUM_ct_md_totalpasivo_T21_nuevo40_7_2	18,9388134
148	SUM_ct_md_totalpasivo_T21_nuevo40_7_3	5,75047353
149	SUM_ct_md_totalpasivo_T21_nuevoperdido	-11,2351427
150	SUM_ct_md_totalpasivo_T24_nuevo40_8_1	48,4341208
151	SUM_ct_md_totalpasivo_T24_nuevo40_8_2	47,5212515
152	SUM_ct_md_totalpasivo_T24_nuevo40_8_3	-12,6628806
153	SUM_ct_md_totalpasivo_T24_nuevoperdido	-4,77076475
154	SUM_ct_md_totalpasivo_T27_nuevo40_9_1	24,4571137
155	SUM_ct_md_totalpasivo_T27_nuevo40_9_2	3,32723938
156	SUM_ct_md_totalpasivo_T27_nuevo40_9_3	10,8343411
157	SUM_ct_md_totalpasivo_T27_nuevoperdido	-0,86066457
158	SUM_ct_md_totalpasivo_T30_nuevo40_10_1	31,1047243
159	SUM_ct_md_totalpasivo_T30_nuevo40_10_2	31,1000246
160	SUM_ct_md_totalpasivo_T30_nuevo40_10_3	-16,2559246
161	SUM_ct_md_totalpasivo_T30_nuevoperdido	-11,4609591
162	SUM_ct_md_totalahorrosaldopromedio_T3_nuevo41_1_2	149,927894
163	SUM_ct_md_totalahorrosaldopromedio_T3_nuevo41_1_3	190,11943
164	SUM_ct_md_totalahorrosaldopromedio_T3_nuevo41_1_4	225,385781
165	SUM_ct_md_totalahorrosaldopromedio_T3_nuevo41_1_5	215,708358
166	SUM_ct_md_totalahorrosaldopromedio_T3_nuevo41_1_6	138,582882
167	SUM_ct_md_totalahorrosaldopromedio_T3_nuevoperdido	-36,903632
168	SUM_ct_md_totalahorrosaldopromedio_T6_nuevo41_2_2	-81,8224529
169	SUM_ct_md_totalahorrosaldopromedio_T6_nuevo41_2_3	-10,7225699
170	SUM_ct_md_totalahorrosaldopromedio_T6_nuevo41_2_4	62,1804339
171	SUM_ct_md_totalahorrosaldopromedio_T6_nuevo41_2_5	101,159822
172	SUM_ct_md_totalahorrosaldopromedio_T6_nuevoperdido	0,61294191
173	SUM_ct_md_totalahorrosaldopromedio_T9_nuevo41_3_2	56,1655403
174	SUM_ct_md_totalahorrosaldopromedio_T9_nuevo41_3_3	24,3168195
175	SUM_ct_md_totalahorrosaldopromedio_T9_nuevo41_3_4	0,87564895
176	SUM_ct_md_totalahorrosaldopromedio_T9_nuevo41_3_5	14,7308101
177	SUM_ct_md_totalahorrosaldopromedio_T9_nuevoperdido	-10,8640684
178	SUM_ct_md_totalahorrosaldopromedio_T12_nuevo41_4_2	0,33750825
179	SUM_ct_md_totalahorrosaldopromedio_T12_nuevo41_4_3	-58,942502
180	SUM_ct_md_totalahorrosaldopromedio_T12_nuevo41_4_4	56,2133424
181	SUM_ct_md_totalahorrosaldopromedio_T12_nuevo41_4_5	7,86263057
182	SUM_ct_md_totalahorrosaldopromedio_T12_nuevoperdido	1,81872946
183	SUM_ct_md_totalahorrosaldopromedio_T15_nuevo41_5_2	-4,60437935
184	SUM_ct_md_totalahorrosaldopromedio_T15_nuevo41_5_3	14,1648909

Cuadro D.5: Pesos del hiperplano de separación, tabla 5.

186	SUM_ct_md_totalahorrosaldopromedio_T15_nuevo41_5_5	-10,9521084
187	SUM_ct_md_totalahorrosaldopromedio_T15_nuevoperdido	-4,45044379
188	SUM_ct_md_totalahorrosaldopromedio_T18_nuevo41_6_1	-15,7560671
189	SUM_ct_md_totalahorrosaldopromedio_T18_nuevo41_6_2	14,4090043
190	SUM_ct_md_totalahorrosaldopromedio_T18_nuevo41_6_3	-13,2684944
191	SUM_ct_md_totalahorrosaldopromedio_T18_nuevoperdido	3,70197949
192	SUM_ct_md_totalahorrosaldopromedio_T21_nuevo41_7_1	-53,8671997
193	SUM_ct_md_totalahorrosaldopromedio_T21_nuevo41_7_2	39,1042375
194	SUM_ct_md_totalahorrosaldopromedio_T21_nuevo41_7_3	32,8996883
195	SUM_ct_md_totalahorrosaldopromedio_T21_nuevoperdido	-11,2351427
196	SUM_ct_md_totalahorrosaldopromedio_T24_nuevo41_8_2	48,8960648
197	SUM_ct_md_totalahorrosaldopromedio_T24_nuevo41_8_3	-3,86458116
198	SUM_ct_md_totalahorrosaldopromedio_T24_nuevoperdido	-4,77076475
199	SUM_ct_md_totalahorrosaldopromedio_T27_nuevo41_9_2	25,0251128
200	SUM_ct_md_totalahorrosaldopromedio_T27_nuevo41_9_3	4,64542852
201	SUM_ct_md_totalahorrosaldopromedio_T27_nuevoperdido	-0,86066457
202	SUM_ct_md_totalahorrosaldopromedio_T30_nuevo41_10_2	17,7427929
203	SUM_ct_md_totalahorrosaldopromedio_T30_nuevo41_10_3	2,61026793
204	SUM_ct_md_totalahorrosaldopromedio_T30_nuevoperdido	-11,4609591
205	SUM_ct_md_totalcorrientesaldopromedio_T3_nuevo42_1_2,	-132,333497
206	SUM_ct_md_totalcorrientesaldopromedio_T3_nuevoperdido	-36,903632
207	SUM_ct_md_totalcorrientesaldopromedio_T6_nuevo42_2_2,	-62,3334972
208	SUM_ct_md_totalcorrientesaldopromedio_T6_nuevoperdido	0,61294191
209	SUM_ct_md_totalcorrientesaldopromedio_T9_nuevo42_3_2,	-40,6174751
210	SUM_ct_md_totalcorrientesaldopromedio_T9_nuevoperdido	-10,8640684
211	SUM_ct_md_totalcorrientesaldopromedio_T12_nuevo42_4_2,	-19,9071272
212	SUM_ct_md_totalcorrientesaldopromedio_T12_nuevoperdido	1,81872946
213	SUM_ct_md_totalcorrientesaldopromedio_T15_nuevo42_5_2,	-26,1069969
214	SUM_ct_md_totalcorrientesaldopromedio_T15_nuevoperdido	-4,45044379
215	SUM_ct_md_totalcorrientesaldopromedio_T18_nuevo42_6_2,	-19,8994685
216	SUM_ct_md_totalcorrientesaldopromedio_T18_nuevoperdido	3,70197949
217	SUM_ct_md_totalcorrientesaldopromedio_T21_nuevo42_7_2,	-19,7906509
218	SUM_ct_md_totalcorrientesaldopromedio_T21_nuevoperdido	-11,2351427
219	SUM_ct_md_totalcorrientesaldopromedio_T24_nuevo42_8_2,	-5,50099876
220	SUM_ct_md_totalcorrientesaldopromedio_T24_nuevoperdido	-4,77076475
221	SUM_ct_md_totalcorrientesaldopromedio_T27_nuevo42_9_2,	8,03616389
222	SUM_ct_md_totalcorrientesaldopromedio_T27_nuevoperdido	-0,86066457
223	SUM_ct_md_totalcorrientesaldopromedio_T30_nuevo42_10_2,	13,0361639
224	SUM_ct_md_totalcorrientesaldopromedio_T30_nuevoperdido	-11,4609591
225	SUM_ct_md_carteraenriesgocredito_T3_nuevo43_1_2,	-19,1151084
226	SUM_ct_md_carteraenriesgocredito_T3_nuevoperdido	-36,903632
227	SUM_ct_md_carteraenriesgocredito_T6_nuevo43_2_2,	-5,06078114
228	SUM_ct_md_carteraenriesgocredito_T6_nuevoperdido	0,61294191
229	SUM_ct_md_carteraenriesgocredito_T9_nuevo43_3_2,	-5,06078114
230	SUM_ct_md_carteraenriesgocredito_T9_nuevoperdido	-10,8640684
231	SUM_ct_md_carteraenriesgocredito_T12_nuevo43_4_2,	-10,7158369
232	SUM_ct_md_carteraenriesgocredito_T12_nuevoperdido	1,81872946
233	SUM_ct_md_carteraenriesgocredito_T15_nuevo43_5_2,	-6,24177479
234	SUM_ct_md_carteraenriesgocredito_T15_nuevoperdido	-4,45044379

Cuadro D.6: Pesos del hiperplano de separación, tabla 6.

236	SUM_ct_md_carteraenriesgocredito_T18_nuevoperdido	3,70197949
237	SUM_ct_md_carteraenriesgocredito_T21_nuevo43_7_2,	2,92206057
238	SUM_ct_md_carteraenriesgocredito_T21_nuevoperdido	-11,2351427
239	SUM_ct_md_carteraenriesgocredito_T24_nuevo43_8_2,	28,1294794
240	SUM_ct_md_carteraenriesgocredito_T24_nuevoperdido	-4,77076475
241	SUM_ct_md_carteraenriesgocredito_T27_nuevo43_9_2,	24,9730226
242	SUM_ct_md_carteraenriesgocredito_T27_nuevoperdido	-0,86066457
243	SUM_ct_md_carteraenriesgocredito_T30_nuevo43_10_2,	9,46465988
244	SUM_ct_md_carteraenriesgocredito_T30_nuevoperdido	-11,4609591
245	SUM_ct_md_totalahorrosaldocontable_T3_nuevo45_1_2	226,688826
246	SUM_ct_md_totalahorrosaldocontable_T3_nuevo45_1_3	230,433953
247	SUM_ct_md_totalahorrosaldocontable_T3_nuevo45_1_4	247,409742
248	SUM_ct_md_totalahorrosaldocontable_T3_nuevo45_1_5	247,942347
249	SUM_ct_md_totalahorrosaldocontable_T3_nuevoperdido	-36,903632
250	SUM_ct_md_totalahorrosaldocontable_T6_nuevo45_2_2	10,1398143
251	SUM_ct_md_totalahorrosaldocontable_T6_nuevo45_2_3	89,6399724
252	SUM_ct_md_totalahorrosaldocontable_T6_nuevo45_2_4	106,015447
253	SUM_ct_md_totalahorrosaldocontable_T6_nuevoperdido	0,61294191
254	SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_2	2,69533408
255	SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_3	-31,3192503
256	SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_4	-19,6844416
257	SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_5	50,7976608
258	SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_6	14,5692686
259	SUM_ct_md_totalahorrosaldocontable_T9_nuevoperdido	-10,8640684
260	SUM_ct_md_totalahorrosaldocontable_T12_nuevo45_4_2	-28,1541616
261	SUM_ct_md_totalahorrosaldocontable_T12_nuevo45_4_3	-34,44787
262	SUM_ct_md_totalahorrosaldocontable_T12_nuevo45_4_4	25,3771721
263	SUM_ct_md_totalahorrosaldocontable_T12_nuevo45_4_5	27,6958388
264	SUM_ct_md_totalahorrosaldocontable_T12_nuevoperdido	1,81872946
265	SUM_ct_md_totalahorrosaldocontable_T15_nuevo45_5_2	34,7598265
266	SUM_ct_md_totalahorrosaldocontable_T15_nuevo45_5_3	-51,4696616
267	SUM_ct_md_totalahorrosaldocontable_T15_nuevo45_5_4	-5,83263194
268	SUM_ct_md_totalahorrosaldocontable_T15_nuevoperdido	-4,45044379
269	SUM_ct_md_totalahorrosaldocontable_T18_nuevo45_6_2	23,0753352
270	SUM_ct_md_totalahorrosaldocontable_T18_nuevo45_6_3	-10,1765696
271	SUM_ct_md_totalahorrosaldocontable_T18_nuevo45_6_4	-5,84467798
272	SUM_ct_md_totalahorrosaldocontable_T18_nuevoperdido	3,70197949
273	SUM_ct_md_totalahorrosaldocontable_T21_nuevo45_7_2	-32,3775406
274	SUM_ct_md_totalahorrosaldocontable_T21_nuevo45_7_3	15,1371403
275	SUM_ct_md_totalahorrosaldocontable_T21_nuevo45_7_4	28,5678815
276	SUM_ct_md_totalahorrosaldocontable_T21_nuevo45_7_5	43,7748612
277	SUM_ct_md_totalahorrosaldocontable_T21_nuevoperdido	-11,2351427
278	SUM_ct_md_totalahorrosaldocontable_T24_nuevo45_8_2	27,5212515
279	SUM_ct_md_totalahorrosaldocontable_T24_nuevo45_8_3	17,5102322
280	SUM_ct_md_totalahorrosaldocontable_T24_nuevoperdido	-4,77076475
281	SUM_ct_md_totalahorrosaldocontable_T27_nuevo45_9_2	7,10025581
282	SUM_ct_md_totalahorrosaldocontable_T27_nuevo45_9_3	27,5702855
283	SUM_ct_md_totalahorrosaldocontable_T27_nuevoperdido	-0,86066457
284	SUM_ct_md_totalahorrosaldocontable_T30_nuevo45_10_2	2,75167324

Cuadro D.7: Pesos del hiperplano de separación, tabla 7.

286	SUM_ct_md_totalahorrosaldocontable_T30_nuevoperdido	-11,4609591
287	SUM_ct_md_totalcorrientesaldocontable_T3_nuevo46_1_2,	-111,755008
288	SUM_ct_md_totalcorrientesaldocontable_T3_nuevoperdido	-36,903632
289	SUM_ct_md_totalcorrientesaldocontable_T6_nuevo46_2_2,	-49,5166663
290	SUM_ct_md_totalcorrientesaldocontable_T6_nuevoperdido	0,61294191
291	SUM_ct_md_totalcorrientesaldocontable_T9_nuevo46_3_2,	-2,80064421
292	SUM_ct_md_totalcorrientesaldocontable_T9_nuevoperdido	-10,8640684
293	SUM_ct_md_totalcorrientesaldocontable_T12_nuevo46_4_2,	43,1826553
294	SUM_ct_md_totalcorrientesaldocontable_T12_nuevoperdido	1,81872946
295	SUM_ct_md_totalcorrientesaldocontable_T15_nuevo46_5_2,	12,9752364
296	SUM_ct_md_totalcorrientesaldocontable_T15_nuevoperdido	-4,45044379
297	SUM_ct_md_totalcorrientesaldocontable_T18_nuevo46_6_2,	14,1827648
298	SUM_ct_md_totalcorrientesaldocontable_T18_nuevoperdido	3,70197949
299	SUM_ct_md_totalcorrientesaldocontable_T21_nuevo46_7_2,	-4,79065091
300	SUM_ct_md_totalcorrientesaldocontable_T21_nuevoperdido	-11,2351427
301	SUM_ct_md_totalcorrientesaldocontable_T24_nuevo46_8_2,	9,49900124
302	SUM_ct_md_totalcorrientesaldocontable_T24_nuevoperdido	-4,77076475
303	SUM_ct_md_totalcorrientesaldocontable_T27_nuevo46_9_2,	23,0361639
304	SUM_ct_md_totalcorrientesaldocontable_T27_nuevoperdido	-0,86066457
305	SUM_ct_md_totalcorrientesaldocontable_T30_nuevo46_10_2,	18,0361639
306	SUM_ct_md_totalcorrientesaldocontable_T30_nuevoperdido	-11,4609591
307	SUM_ct_md_totaldepplazosaldocontable_T3_nuevo47_1_2,	-138,91396
308	SUM_ct_md_totaldepplazosaldocontable_T3_nuevoperdido	-36,903632
309	SUM_ct_md_totaldepplazosaldocontable_T6_nuevo47_2_2,	12,9759998
310	SUM_ct_md_totaldepplazosaldocontable_T6_nuevoperdido	0,61294191
311	SUM_ct_md_totaldepplazosaldocontable_T9_nuevo47_3_2,	-11,3669513
312	SUM_ct_md_totaldepplazosaldocontable_T9_nuevoperdido	-10,8640684
313	SUM_ct_md_totaldepplazosaldocontable_T12_nuevo47_4_2,	18,5242311
314	SUM_ct_md_totaldepplazosaldocontable_T12_nuevoperdido	1,81872946
315	SUM_ct_md_totaldepplazosaldocontable_T15_nuevo47_5_2,	3,5242311
316	SUM_ct_md_totaldepplazosaldocontable_T15_nuevoperdido	-4,45044379
317	SUM_ct_md_totaldepplazosaldocontable_T18_nuevo47_6_2,	-22,364077
318	SUM_ct_md_totaldepplazosaldocontable_T18_nuevoperdido	3,70197949
319	SUM_ct_md_totaldepplazosaldocontable_T21_nuevo47_7_2,	-57,2552594
320	SUM_ct_md_totaldepplazosaldocontable_T21_nuevoperdido	-11,2351427
321	SUM_ct_md_totaldepplazosaldocontable_T24_nuevo47_8_2,	-36,7821187
322	SUM_ct_md_totaldepplazosaldocontable_T24_nuevoperdido	-4,77076475
323	SUM_ct_md_totaldepplazosaldocontable_T27_nuevo47_9_2,	-32,3200469
324	SUM_ct_md_totaldepplazosaldocontable_T27_nuevoperdido	-0,86066457
325	SUM_ct_md_totaldepplazosaldocontable_T30_nuevo47_10_2,	-14,0498826
326	SUM_ct_md_totaldepplazosaldocontable_T30_nuevoperdido	-11,4609591
327	SUM_ct_md_carteraenriesgotarjeta_T12_nuevo44_4_2,	-42,9736081
328	SUM_ct_md_carteraenriesgotarjeta_T12_nuevoperdido	1,81872946
329	SUM_ct_md_carteraenriesgotarjeta_T15_nuevo44_5_2,	-42,9736081
330	SUM_ct_md_carteraenriesgotarjeta_T15_nuevoperdido	-4,45044379
331	SUM_ct_md_carteraenriesgotarjeta_T18_nuevo44_6_2,	-42,9736081
332	SUM_ct_md_carteraenriesgotarjeta_T18_nuevoperdido	3,70197949
333	SUM_ct_md_carteraenriesgotarjeta_T21_nuevo44_7_2,	-42,9736081
334	SUM_ct_md_carteraenriesgotarjeta_T21_nuevoperdido	-11,2351427

Cuadro D.8: Pesos del hiperplano de separación, tabla 8.

336	SUM_ct_md_carteraenriesgotarjeta_T24_nuevoperdido	-4,77076475
337	SUM_ct_md_carteratotalcredito_T3_nuevo48_1_2	21,2894809
338	SUM_ct_md_carteratotalcredito_T3_nuevo48_1_3	-3,43457484
339	SUM_ct_md_carteratotalcredito_T3_nuevo48_1_4	-5,57147278
340	SUM_ct_md_carteratotalcredito_T3_nuevoperdido	-36,903632
341	SUM_ct_md_carteratotalcredito_T6_nuevo48_2_2	22,3941201
342	SUM_ct_md_carteratotalcredito_T6_nuevo48_2_3	5,06648509
343	SUM_ct_md_carteratotalcredito_T6_nuevo48_2_4	11,3016951
344	SUM_ct_md_carteratotalcredito_T6_nuevoperdido	0,61294191
345	SUM_ct_md_carteratotalcredito_T9_nuevo48_3_2	-5,87969761
346	SUM_ct_md_carteratotalcredito_T9_nuevo48_3_3	14,4250457
347	SUM_ct_md_carteratotalcredito_T9_nuevo48_3_4	30,2169522
348	SUM_ct_md_carteratotalcredito_T9_nuevoperdido	0,61294191
349	SUM_ct_md_carteratotalcredito_T12_nuevo48_4_2	-3,73401455
350	SUM_ct_md_carteratotalcredito_T12_nuevo48_4_3	13,2160274
351	SUM_ct_md_carteratotalcredito_T12_nuevo48_4_4	17,0712911
352	SUM_ct_md_carteratotalcredito_T12_nuevoperdido	1,81872946
353	SUM_ct_md_carteratotalcredito_T15_nuevo48_5_2	8,65116856
354	SUM_ct_md_carteratotalcredito_T15_nuevo48_5_3	-13,5998455
355	SUM_ct_md_carteratotalcredito_T15_nuevo48_5_4	-42,3554044
356	SUM_ct_md_carteratotalcredito_T15_nuevoperdido	-4,45044379
357	SUM_ct_md_carteratotalcredito_T18_nuevo48_6_2	-33,5619678
358	SUM_ct_md_carteratotalcredito_T18_nuevo48_6_3	-5,3108473
359	SUM_ct_md_carteratotalcredito_T18_nuevo48_6_4	-27,2237379
360	SUM_ct_md_carteratotalcredito_T18_nuevoperdido	3,70197949
361	SUM_ct_md_carteratotalcredito_T21_nuevo48_7_2	-28,5619678
362	SUM_ct_md_carteratotalcredito_T21_nuevo48_7_3	7,52925157
363	SUM_ct_md_carteratotalcredito_T21_nuevo48_7_4	-10,3903896
364	SUM_ct_md_carteratotalcredito_T21_nuevoperdido	-11,2351427
365	SUM_ct_md_carteratotalcredito_T24_nuevo48_8_2	-5,27083432
366	SUM_ct_md_carteratotalcredito_T24_nuevo48_8_3	-5,10527316
367	SUM_ct_md_carteratotalcredito_T24_nuevo48_8_4	-16,7573462
368	SUM_ct_md_carteratotalcredito_T24_nuevoperdido	-4,77076475
369	SUM_ct_md_carteratotalcredito_T27_nuevo48_9_2	7,80445562
370	SUM_ct_md_carteratotalcredito_T27_nuevo48_9_3	-10,7365222
371	SUM_ct_md_carteratotalcredito_T27_nuevo48_9_4	0,35876835
372	SUM_ct_md_carteratotalcredito_T27_nuevoperdido	-0,86066457
373	SUM_ct_md_carteratotalcredito_T30_nuevo48_9_2	-21,1233031
374	SUM_ct_md_carteratotalcredito_T30_nuevo48_9_4	0,1976156
375	SUM_ct_md_carteratotalcredito_T30_nuevoperdido	-11,4609591
376	SUM_ct_md_carteratotaltarjeta_T3_nuevo49_1_2,	11,7606447
377	SUM_ct_md_carteratotaltarjeta_T3_nuevoperdido	-36,903632
378	SUM_ct_md_carteratotaltarjeta_T6_nuevo49_2_2,	6,7606447
379	SUM_ct_md_carteratotaltarjeta_T6_nuevoperdido	0,61294191
380	SUM_ct_md_carteratotaltarjeta_T9_nuevo49_3_2,	1,7606447
381	SUM_ct_md_carteratotaltarjeta_T9_nuevoperdido	-10,8640684
382	SUM_ct_md_carteratotaltarjeta_T12_nuevo49_4_2,	6,7606447
383	SUM_ct_md_carteratotaltarjeta_T12_nuevoperdido	1,81872946
384	SUM_ct_md_carteratotaltarjeta_T15_nuevo49_5_2,	1,7606447

Cuadro D.9: Pesos del hiperplano de separación, tabla 9.

386	SUM_ct_md_carteratotaltarjeta_T18_nuevo49_6_2,	1,7606447
387	SUM_ct_md_carteratotaltarjeta_T18_nuevoperdido	3,70197949
388	SUM_ct_md_carteratotaltarjeta_T21_nuevo49_7_2,	-3,2393553
389	SUM_ct_md_carteratotaltarjeta_T21_nuevoperdido	-11,2351427
390	SUM_ct_md_carteratotaltarjeta_T24_nuevo49_8_2,	6,7606447
391	SUM_ct_md_carteratotaltarjeta_T24_nuevoperdido	-4,77076475
392	SUM_ct_md_carteratotaltarjeta_T27_nuevo49_9_2,	21,7606447
393	SUM_ct_md_carteratotaltarjeta_T27_nuevoperdido	-0,86066457
394	SUM_ct_md_carteratotaltarjeta_T30_nuevo49_10_2,	9,73425277
395	SUM_ct_md_carteratotaltarjeta_T30_nuevoperdido	-11,4609591
396	SUM_ct_md_numeroperaciones_activo_T3_nuevo50_1_2	13,9855007
397	SUM_ct_md_numeroperaciones_activo_T3_nuevo50_1_3	-76,9509178
398	SUM_ct_md_numeroperaciones_activo_T3_nuevo50_1_4	2,41570047
399	SUM_ct_md_numeroperaciones_activo_T3_nuevoperdido	-7,04787517
400	SUM_ct_md_numeroperaciones_activo_T6_nuevo50_2_2	-16,5219443
401	SUM_ct_md_numeroperaciones_activo_T6_nuevo50_2_3	-2,58429954
402	SUM_ct_md_numeroperaciones_activo_T6_nuevoperdido	0,61294191
403	SUM_ct_md_numeroperaciones_activo_T9_nuevo50_3_2	-11,5219443
404	SUM_ct_md_numeroperaciones_activo_T9_nuevo50_3_3	-13,2393553
405	SUM_ct_md_numeroperaciones_activo_T9_nuevoperdido	-10,8640684
406	SUM_ct_md_numeroperaciones_activo_T12_nuevo50_4_2	-1,12211271
407	SUM_ct_md_numeroperaciones_activo_T12_nuevo50_4_3	-9,75641071
408	SUM_ct_md_numeroperaciones_activo_T12_nuevoperdido	1,81872946
409	SUM_ct_md_numeroperaciones_activo_T15_nuevo50_5_2	-35,9146939
410	SUM_ct_md_numeroperaciones_activo_T15_nuevo50_5_3	-4,75641071
411	SUM_ct_md_numeroperaciones_activo_T15_nuevoperdido	-4,45044379
412	SUM_ct_md_numeroperaciones_activo_T18_nuevo50_6_2	-17,7477857
413	SUM_ct_md_numeroperaciones_activo_T18_nuevo50_6_3	-50,9146939
414	SUM_ct_md_numeroperaciones_activo_T18_nuevo50_6_4	5,24358929
415	SUM_ct_md_numeroperaciones_activo_T18_nuevoperdido	3,70197949
416	SUM_ct_md_numeroperaciones_activo_T21_nuevo50_7_2	1,92566143
417	SUM_ct_md_numeroperaciones_activo_T21_nuevo50_7_3	-35,9146939
418	SUM_ct_md_numeroperaciones_activo_T21_nuevo50_7_4	10,2435893
419	SUM_ct_md_numeroperaciones_activo_T21_nuevoperdido	-11,2351427
420	SUM_ct_md_numeroperaciones_activo_T24_nuevo50_8_2	11,3501658
421	SUM_ct_md_numeroperaciones_activo_T24_nuevo50_8_3	-11,0495461
422	SUM_ct_md_numeroperaciones_activo_T24_nuevo50_8_4	5,24358929
423	SUM_ct_md_numeroperaciones_activo_T24_nuevoperdido	-4,77076475
424	SUM_ct_md_numeroperaciones_activo_T27_nuevo50_9_2	22,5649912
425	SUM_ct_md_numeroperaciones_activo_T27_nuevo50_9_3	-2,70421594
426	SUM_ct_md_numeroperaciones_activo_T27_nuevo50_9_4	10,2435893
427	SUM_ct_md_numeroperaciones_activo_T27_nuevoperdido	-0,86066457
428	SUM_ct_md_numeroperaciones_activo_T30_nuevo50_10_2	-22,4350088
429	SUM_ct_md_numeroperaciones_activo_T30_nuevo50_10_3	34,8651478
430	SUM_ct_md_numeroperaciones_activo_T30_nuevo50_10_4	-10,6781637
431	SUM_ct_md_numeroperaciones_activo_T30_nuevoperdido	-11,4609591
432	SUM_ct_cod_marca_productocartera_T3_nuevo51_1_2	48,8412575
433	SUM_ct_cod_marca_productocartera_T3_nuevo51_1_3	-79,5352173
434	SUM_ct_cod_marca_productocartera_T3_nuevoperdido	-36,903632

Cuadro D.10: Add caption

436	SUM_ct_cod_marca_productocartera_T6_nuevoperdido	0,61294191
437	SUM_ct_cod_marca_productocartera_T9_nuevo51_3_2	-29,4955524
438	SUM_ct_cod_marca_productocartera_T9_nuevoperdido	-10,8640684
439	SUM_ct_cod_marca_productocartera_T12_nuevo51_4_2	-24,0957208
440	SUM_ct_cod_marca_productocartera_T12_nuevoperdido	1,81872946
441	SUM_ct_cod_marca_productocartera_T15_nuevo51_5_2	-58,8883019
442	SUM_ct_cod_marca_productocartera_T15_nuevoperdido	-4,45044379
443	SUM_ct_cod_marca_productocartera_T18_nuevo51_6_2	-58,8883019
444	SUM_ct_cod_marca_productocartera_T18_nuevoperdido	3,70197949
445	SUM_ct_cod_marca_productocartera_T21_nuevo51_7_2	-38,8883019
446	SUM_ct_cod_marca_productocartera_T21_nuevoperdido	-11,2351427
447	SUM_ct_cod_marca_productocartera_T24_nuevo51_8_2	-4,77758106
448	SUM_ct_cod_marca_productocartera_T24_nuevo51_8_3	46,8877547
449	SUM_ct_cod_marca_productocartera_T24_nuevo51_8_4	-36,5659646
450	SUM_ct_cod_marca_productocartera_T24_nuevoperdido	-4,77076475
451	SUM_ct_cod_marca_productocartera_T27_nuevo51_9_2	-3,56275571
452	SUM_ct_cod_marca_productocartera_T27_nuevo51_9_3	49,210092
453	SUM_ct_cod_marca_productocartera_T27_nuevo51_9_4	-15,5429718
454	SUM_ct_cod_marca_productocartera_T27_nuevoperdido	-0,86066457
455	SUM_ct_cod_marca_productocartera_T30_nuevo51_10_2	-21,2404184
456	SUM_ct_cod_marca_productocartera_T30_nuevo51_10_3	66,8877547
457	SUM_ct_cod_marca_productocartera_T30_nuevo51_10_4	-43,8953611
458	SUM_ct_cod_marca_productocartera_T30_nuevoperdido	-11,4609591
459	coeficiente	0,5960203

Anexo E

Resultados obtenidos con regresión logística

Cuadro E.1: Resultados con regresión logística. Elaborado por: Autor

	Estimate	**Std. Error**	**z value**	**Pr**$(>\|z\|)$
(Intercept)	-198.436	7.056.146	-0.028	0.97756
SUM_ct_md_carteratotaltarjeta_T30_nuevo49_10_2	-157.136	12.996.463	-0.012	0.99035
SUM_ct_md_carteratotaltarjeta_T30_nuevoperdido	17.564	0.2101	8.358	¡2e-16 ***
SUM_ct_md_totalpasivo_T3_nuevo40_1_2	-31.810	0.3012	-10.563	¡2e-16 ***
SUM_ct_md_totalpasivo_T3_nuevo40_1_3	-24.835	0.3330	-7.458	8.81e-14 ***
SUM_ct_md_totalpasivo_T3_nuevo40_1_4	-19.458	0.2883	-6.750	1.48e-11 ***
SUM_ct_md_totalpasivo_T3_nuevo40_1_5	-0.4333	0.2990	-1.449	0.14730
SUM_ct_md_totalpasivo_T3_nuevoperdido	182.396	7.403.368	0.025	0.98034
SUM_ct_md_carteratotaltarjeta_T12_nuevo49_4_2	-18.830	0.8778	-2.145	0.03194 *
SUM_ct_md_carteratotaltarjeta_T12_nuevoperdido	0.4770	0.2571	1.855	0.06355 .
SUM_ct_md_totalahorrosaldocontable_T6_nuevo45_2_2	-0.4180	0.6933	-0.603	0.54657
SUM_ct_md_totalahorrosaldocontable_T6_nuevo45_2_3	-11.779	0.7685	-1.533	0.12538
SUM_ct_md_totalahorrosaldocontable_T6_nuevo45_2_4	-24.898	0.8803	-2.828	0.00468 **
SUM_ct_md_totalahorrosaldocontable_T6_nuevoperdido	183.489	7.056.148	0.026	0.97925
SUM_ct_md_totalproductoscliente_T6_nuevo38_2_2	206.968	7.056.146	0.029	0.97660
SUM_ct_md_totalproductoscliente_T6_nuevo38_2_3	181.405	7.056.147	0.026	0.97949
SUM_ct_md_totalproductoscliente_T6_nuevo38_2_4	176.482	7.056.147	0.025	0.98005
SUM_ct_md_totalproductoscliente_T6_nuevo38_2_5	195.627	7.056.146	0.028	0.97788
SUM_ct_md_totalproductoscliente_T6_nuevo38_2_6	196.800	7.056.146	0.028	0.97775
SUM_ct_md_totalproductoscliente_T6_nuevoperdido	NA	NA	NA	NA
SUM_ct_md_totalahorrosaldopromedio_T6_nuevo41_2_2	10.678	0.7985	1.337	0.18118
SUM_ct_md_totalahorrosaldopromedio_T6_nuevo41_2_3	0.6913	0.8595	0.804	0.42122
SUM_ct_md_totalahorrosaldopromedio_T6_nuevo41_2_4	0.9693	0.8584	1.129	0.25887
SUM_ct_md_totalahorrosaldopromedio_T6_nuevo41_2_5	0.7317	0.9246	0.791	0.42875
SUM_ct_md_totalahorrosaldopromedio_T6_nuevoperdido	NA	NA	NA	NA
SUM_ct_md_totalahorrosaldopromedio_T9_nuevo41_3_2	-197.186	38.375.752	-0.005	0.99590
SUM_ct_md_totalahorrosaldopromedio_T9_nuevo41_3_3	-203.688	38.375.752	-0.005	0.99577
SUM_ct_md_totalahorrosaldopromedio_T9_nuevo41_3_4	-196.993	38.375.752	-0.005	0.99590
SUM_ct_md_totalahorrosaldopromedio_T9_nuevo41_3_5	-184.073	38.375.752	-0.005	0.99617
SUM_ct_md_totalahorrosaldopromedio_T9_nuevoperdido	0.5877	0.5744	1.023	0.30619
SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_2	184.311	38.375.752	0.005	0.99617
SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_3	188.676	38.375.752	0.005	0.99608
SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_4	187.463	38.375.752	0.005	0.99610
SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_5	181.269	38.375.753	0.005	0.99623
SUM_ct_md_totalahorrosaldocontable_T9_nuevo45_3_6	187.950	38.375.753	0.005	0.99609
SUM_ct_md_totalahorrosaldocontable_T9_nuevoperdido	NA	NA	NA	NA
SUM_ct_md_numeroperaciones_activo_T9_nuevo50_3_2	12.893	0.3118	4.135	3.55e-05 ***
SUM_ct_md_numeroperaciones_activo_T9_nuevo50_3_3	-0.4856	10.742	-0.452	0.65124
SUM_ct_md_numeroperaciones_activo_T9_nuevoperdido	NA	NA	NA	NA

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anexo F

Divergencia Kullback-Leibler

Sea x variable aleatoria discreta, ahora nos preguntamos cuánta información se recibe cuando observamos un valor específico para esta variable. La cantidad de información puede verse como el "grado de sorpresa" al aprender el valor de x. Si nos dicen que acaba de ocurrir un evento altamente improbable, se habrá recibido más información si nos hubieran dicho que algo muy probable acaba de ocurrir, y si supiéramos que el evento sucederá con seguridad no recibiríamos información. La medida de contenido de información, por lo tanto, dependerá de la distribución de probabilidad $p(x)$, y por lo tanto se busca una cantidad $h(x)$ que sea una función monotóna de la probabilidad $p(x)$ y que expresa el contenido de información. La forma de $h(.)$ se puede encontrar al observar dos eventos x e y que no están relacionados, entonces la ganancia de información al observarlos debe ser la suma de la información obtenida de cada uno de ellos por separado, de modo que $h(x, y) = h(x) + h(y)$. Dos eventos no relacionados serán estadísticamente independientes y no relacionados cuando $p(x, y) = p(x)p(y)$. A partir de estas dos relaciones, se muestra fácilmente que $h(x)$ debe estar dado por el logaritmo de $p(x)$. Tenga en cuenta que los eventos de baja probabilidad x corresponden a un alto contenido de información. La elección de la base para el logaritmo es arbitraria.

La teoría de la información define el grado de dependencia estadística entre dos variables aleatorias a partir del concepto de información mutua. El nombre hace referencia al hecho de que lo que se mide es la información que contiene una variable acerca del valor de alguna otra. La información mutua puede definirse a partir de divergencia de Kullback-Leibler (KL), este valor es el método más conocido para la medición de distancias entre dos distribuciones de probabilidad. Esta divergencia viene dada, para dos distribuciones de probabilidades discretas P y Q con el mismo dominio $X, Y \in x_1, x_2, ...x_L$ está dada por:

$$D_{kl}(P(X), Q(X)) = -\sum_{l=1}^{L} P(x_l) log \frac{P(x_l)}{Q(x_l)}.$$

Si $P(X) = Q(X)$, entonces $D_{kl}(P(X), Q(X))$. Caso contrario, $D_{kl}(P(X), Q(X)) > 0$ [Bishop, 2006], [Murillo, 2007].

Anexo G

Comportamiento de un cliente desertor

Los siguientes gráficos indican el comportamiento de un cliente desertor particular cuya probabilidad de deserción fue de $0,93$.
Los valores de *total corrientes saldo contable* para un cliente son el monto total de los saldos promedios en cuentas corrientes del cliente al mes de corte.

Los valores de *total corrientes saldo promedio* para un cliente son el monto total de

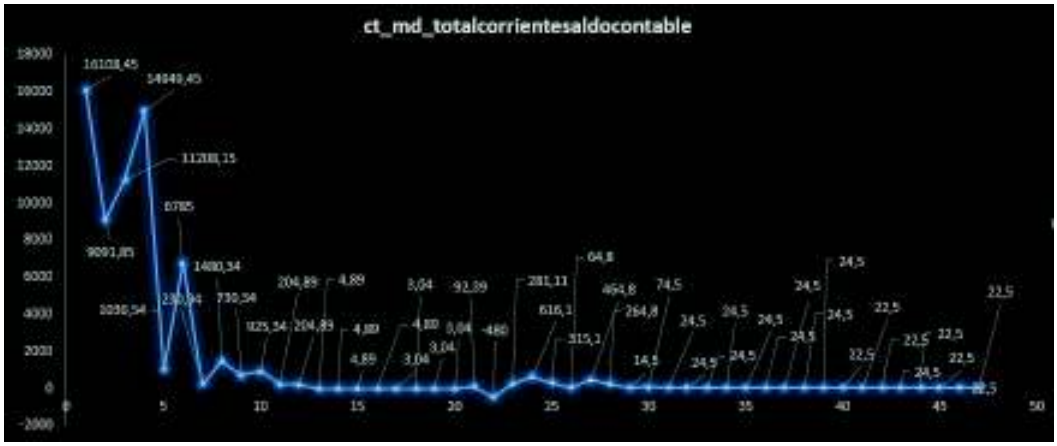

Figura G.1: Total corrientes saldo contable para un cliente desertor. Elaborado por: Autor

los saldos contables en cuentas de ahorro del cliente al mes de corte.

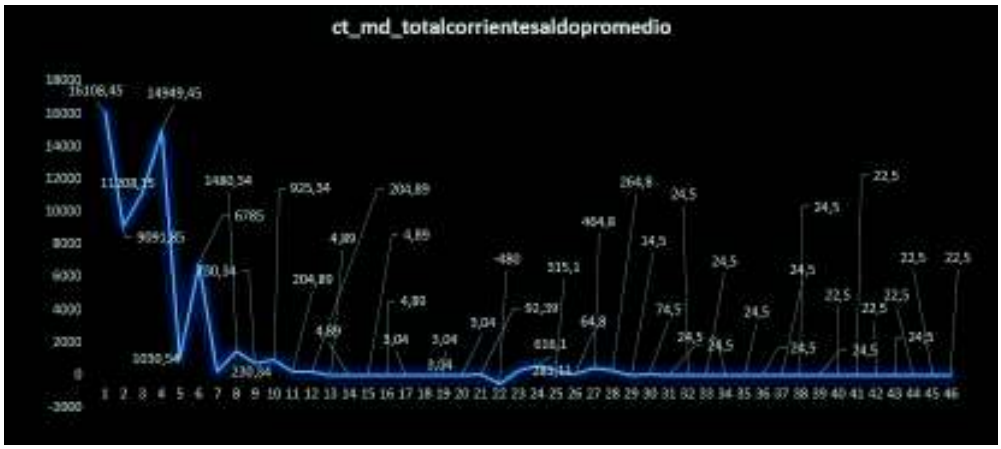

Figura G.2: Total corrientes saldo promedio para un cliente desertor. Elaborado por: Autor

Bibliografía

[Ali and Arıtürk, 2014] Ali, Ö. G. and Arıtürk, U. (2014). Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Systems with Applications*, 41(17):7889–7903.

[Au et al., 2003] Au, T., Ma, G., and Li, S. (2003). Applying and evaluating models to predict customer attrition using data mining techniques. *Journal of Comparative International Management*, 6(1).

[Audzeyeva and Schenk-Hopp, 2012] Audzeyeva, A., S. B. and Schenk-Hopp, K. R. (2012). Forecasting customer behaviour in a multi-service financial organisation: A profitability perspective. *International journal of forecasting*, 28(2):507–518.

[Barga et al., 2015] Barga, R., Fontama, V., Tok, W. H., and Cabrera-Cordon, L. (2015). *Predictive analytics with Microsoft Azure machine learning*. Springer.

[Bhambri, 2013] Bhambri, V. (2013). Data mining as a tool to predict churn behaviour of customers. *International Journal of Management Research*, pages 59–69.

[Bijmolt and Saffert, 2010] Bijmolt, T. H., L. P. S. B. F. E. M. H. B. G. L. A. and Saffert, P. (2010). Analytics for customer engagement. *Journal of Service Research*, 13(3):341–356.

[Bilgen,] Bilgen, O. Churn management of electronic banking customers.

[Birtolo and Ritrovato, 2013] Birtolo, C., D. V. D. C. D. and Ritrovato, P. (2013). Customer churn detection system: Identifying customers who wish to leave a merchant. pages 411–420.

[Bishop, 2006] Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

[Blattberg and Neslin, 2008] Blattberg, R. C. B.-D. K. and Neslin, S. A. (2008). Database marketing: Analyzing and managing customers. international series in quantitative marketing. *International Series in Quantitative Marketing*.

[Burez and Van den Poel, 2009] Burez, J. and Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3):4626–4636.

[Burges, 1998] Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167.

[Buttle, 2009] Buttle, F. (2009). Customer relationship management: concepts and technologies. 5.

[Chandar and Krishna, 2006] Chandar, M., L. A. and Krishna, P. (2006). Modeling churn behavior of bank customers using predictive data mining techniques. pages 24–26.

[Chiang and Lin, 2003] Chiang, D. A., W.-Y. F. L. S. L. and Lin, C. J. (2003). Goal-oriented sequential pattern for network banking churn analysis. *Expert Systems with Applications*, 25(3):293–302.

[Chih et al., 2003] Chih, W., Chih, C., and Chih, J. (2003). A practical guide to support vector classification. *Taipei: Department of Computer Science, National Taiwan University*.

[Chitra and Subashini, 2011] Chitra, K. and Subashini, B. (2011). Customer retention in banking sector using predictive data mining technique.

[Coussement and Van den Poel, 2008] Coussement, K. and Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert systems with applications*, 34(1):313–327.

[Dash and Liu, 1997] Dash, M. and Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, 1(1-4):131–156.

[Dawes and Swailes, 1999] Dawes, J. and Swailes, S. (1999). Retention sans frontieres: issues for financial service retailers. *International Journal of Bank Marketing*, 17(1):36–43.

[Donkers and de Jong, 2007] Donkers, B., V. P. C. and de Jong, M. G. (2007). Modeling clv: A test of competing models in the insurance industry. *Quantitative Marketing and Economics*, 5(2):163–190.

[Esposito, 2005] Esposito, F. (2005). *Innovations in Applied Artificial Intelligence*. Springer.

[Fienberg and Casella,] Fienberg, S. and Casella, G. Springer texts in statistics.

[Garcia Gomez, 2014] Garcia Gomez, D. (2014). Exploration of customer churn routes using machine learning probabilistic models.

[Garland, 2003] Garland, R. (2003). Investigating indicators of customer profitability in personal retail banking. pages 18–21.

[Grosan and Abraham, 2011] Grosan, C. and Abraham, A. (2011). *Intelligent systems*. Springer.

[Hadden et al., 2007] Hadden, J., Tiwari, A., Roy, R., and Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10):2902–2917.

[He and Zhao, 2014] He, B., S. Y. W. Q. and Zhao, X. (2014). Prediction of customer attrition of commercial banks based on svm model. *Procedia Computer Science*, 31:423–430.

[Hernandez and Gallegos, 2004] Hernandez, D. P. and Gallegos, I. Q. (2004). Diseno de una estrategia tecnologica de customer relationship management (crm) para la empresa bpm de mexico.

[Hsu et al., 2003] Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al. (2003). A practical guide to support vector classification.

[Hu, 2005] Hu, X. (2005). A data mining approach for retailing bank customer attrition analysis. *Applied Intelligence*, 22(1):47–60.

[Hwang and Suh, 2004] Hwang, H., J. T. and Suh, E. (2004). An ltv model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert systems with applications*, 26(2):181–188.

[Jakkula, 2006] Jakkula, V. (2006). Tutorial on support vector machine (svm). *School of EECS, Washington State University*, 37.

[Kamakura and Srivastava, 1991] Kamakura, W. A., R. S. N. and Srivastava, R. K. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing*, 8(4):329–349.

[Kamakura and Mazzon, 2003] Kamakura, W. A., W. M. D. R. F. and Mazzon, J. A. (2003). Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in marketing*, 20(1):45–65.

[Kaur et al., 2013] Kaur, M., Singh, K., and Sharma, N. (2013). Data mining as a tool to predict the churn behaviour among indian bank customers. *Int. J. Recent Innov. Trends Comput. Commun*, 1(9):720–725.

[KhakAbi and Namvar, 2010] KhakAbi, S., G. M. R. and Namvar, M. (2010). Data mining applications in customer churn management. In *Intelligent systems, modelling and simulation (ISMS), 2010 International Conference on*, pages 220–225. IEEE.

[Khanna and Awad, 2015] Khanna, R. and Awad, M. (2015). *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Apress.

[Kotsiantis et al., 2006] Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3):159–190.

[Lemmens and Croux, 2006] Lemmens, A. and Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2):276–286.

[Lew and H.,] Lew, A. and H., M. Introduction to data mining principles, studies in computational intelligence (sci).

[Lima and Baesens, 2009] Lima, E., M. C. and Baesens, B. (2009). Domain knowledge integration in data mining using decision tables: Case studies in churn prediction. *Operational Research Society*, 60(8):1096–1106.

[Meyer-Waarden, 2007] Meyer-Waarden, L. (2007). The effects of loyalty programs on customer lifetime duration and share of wallet. *Journal of Retailing*, 83(2):223–236.

[Migueis and e Cunha, 2012] Migueis, V. L., V. d. P. D. C. A. S. and e Cunha, J. F. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert systems with applications*, 39(12):11250–11256.

[Mihelis et al., 2001] Mihelis, G., Grigoroudis, E., Siskos, Y., Politis, Y., and Malandrakis, Y. (2001). Customer satisfaction measurement in the private bank sector. *European Journal of Operational Research*, 130(2):347–360.

[Momparler Pechun, 2008] Momparler Pechun, A. (2008). El desarrollo de la banca electrónica en españa. un análisis comparativo entre entidades online y tradicionales en españa y en estados unidos (doctoral dissertation).

[Morgan and Hunt, 1994] Morgan, R. M. and Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *The journal of marketing*, pages 20–38.

[Morik,] Morik, K. Applications of knowledge discovery.

[Murillo, 2007] Murillo, J. M. L. (2007). *Extracción de características mediante criterios basados en teoría de la información*. PhD thesis, Universidad Carlos III de Madrid.

[Mutanen and Nousiainen, 2006] Mutanen, T., A. J. and Nousiainen, S. (2006). Customer churn prediction-a case study in retail banking. In *Proc. of ECML/PKDD Workshop on Practical Data Mining*, pages 13–19.

[Neslin et al., 2006] Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., and Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of marketing research*, 43(2):204–211.

[Ng and Liu, 2000] Ng, K. and Liu, H. (2000). Customer retention via data mining. *Artificial Intelligence Review*, 14(6):569–590.

[Oliver, 1999] Oliver, R. L. (1999). Whence consumer loyalty? *the Journal of Marketing*, pages 33–44.

[Olkin, 2002] Olkin, G. C. S. F. I. (2002). Springer texts in statistics.

[Osuna et al., 1997] Osuna, E., Freund, R., and Girosi, F. (1997). Support vector machines: Training and applications.

[Peppard, 2000] Peppard, J. (2000). Customer relationship management (crm) in financial services. *European Management Journal*, 18(3):312–327.

[Pontil and Verri, 1998] Pontil, M. and Verri, A. (1998). Properties of support vector machines. *Neural Computation*, 10(4):955–974.

[Popovic and Basic, 2009] Popovic, D. and Basic, B. D. (2009). Churn prediction model in retail banking using fuzzy c-means algorithm. *Informatica*, 33(2).

[Prasad and Madhavi, 2012] Prasad, U. D. and Madhavi, S. (2012). Prediction of churn behavior of bank customers using data mining tools. *Business Intelligence Journal*, 5(1):96–101.

[Reichheld, 1996] Reichheld, F. F. (1996). Learning from customer defections.

[Reichheld and Kenny, 1990] Reichheld, F. F. and Kenny, D. W. (1990). The hidden advantages of customer retention. *Journal of Retail Banking*, 12(4):19–24.

[Rodpysh and Aghai, 2012] Rodpysh, K. V. and Aghai, A. (2012). Applying data mining in customer relationship management. *International Journal of Information Technology Control and Automation (IJITCA)*, 2(3):15–25.

[Seufert, 2013] Seufert, E. B. (2013). *Freemium economics: Leveraging analytics and user segmentation to drive revenue*.

[Sharp, 2002] Sharp, D. E. (2002). Customer relationship management systems handbook. *CRC Press.*

[Shashua, 2009] Shashua, A. (2009). Introduction to machine learning: Class notes 67577. *arXiv preprint arXiv:0904.3664*.

[Singh and Jain, 2010] Singh, S. S. and Jain, D. C. (2010). Measuring customer lifetime value. pages 37–62.

[Stone and Liyanearachchi, 2007] Stone, M. and Liyanearachchi, D. (2007). Managing retention who does it well? *Journal of Database Marketing and Customer Strategy Management*, 14(2):90–103.

[Tang et al., 2009] Tang, Y., Zhang, Y.-Q., Chawla, N. V., and Krasser, S. (2009). Svms modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):281–288.

[Tsiptsis and Chorianopoulos, 2011] Tsiptsis, K. K. and Chorianopoulos, A. (2011). *Data mining techniques in CRM: inside customer segmentation*. John Wiley & Sons.

[Vafeiadis et al., 2015] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., and Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55:1–9.

[Van den Poel and Lariviere, 2004] Van den Poel, D. and Lariviere, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European journal of operational research*, 157(1):196–217.

[Vashishtha, 2015] Vashishtha, S. (2015). Customer retention-fact or fallacy. *GIAN JYOTI E*, 5.

[Weiss, 2004] Weiss, G. M. (2004). Mining with rarity: a unifying framework. *ACM Sigkdd Explorations Newsletter*, 6(1):7–19.

[Witten and Pal, 2016] Witten, I. H., F. E. H. M. A. and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

[Xia and Jin, 2008] Xia, G. E. and Jin, W. D. (2008). Model of customer churn prediction on support vector machine. *Systems Engineering-Theory & Practice*, 28(1):71–77.

[Xiao et al., 2012] Xiao, J., Xie, L., He, C., and Jiang, X. (2012). Dynamic classifier ensemble model for customer classification with imbalanced class distribution. *Expert Systems with Applications*, 39(3):3668–3675.

[Xie and Ying, 2009] Xie, Y., L. X. N. E. W. T. and Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3):5445–5449.

[Yu and Kim, 2012] Yu, H. and Kim, S. (2012). Svm tutorial-classification, regression and ranking. In *Handbook of Natural computing*, pages 479–506. Springer.

[Zhao and Dang, 2008] Zhao, J. and Dang, X. H. (2008). Bank customer churn prediction based on support vector machine: Taking a commercial bank's vip customer churn as the example. In *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, pages 1–4. IEEE.

[Zhao and Ren, 2005] Zhao, Y., L. B. L. X. L. W. and Ren, S. (2005). Customer churn prediction using improved one-class support vector machine. In *International Conference on Advanced Data Mining and Applications*, pages 300–306. Springer.