



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

ESTUDIO DEL COMPORTAMIENTO DE LOS PROVEEDORES DEL ESTADO EN EL SISTEMA CHILECOMPRA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

FRANCISCO JAVIER CORREA QUEZADA

PROFESOR GUÍA:

LUIS ZAVIEZO SCHWARTZMAN

MIEMBROS DE LA COMISIÓN:

CHRISTIAN POTOČNJAK CABRERA.

RICHARD WEBER HAAS

SANTIAGO DE CHILE

SEPTIEMBRE 2010

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL
POR: FRANCISCO JAVIER CORREA QUEZADA
FECHA: 06/09/2010
PROF. GUIA: SR: LUIS ZAVIEZO S.

ESTUDIO DEL COMPORTAMIENTO DE LOS PROVEEDORES DEL ESTADO EN EL SISTEMA CHILECOMPRA

La deserción de clientes de un sistema es un fenómeno que atañe a una gran cantidad de instituciones públicas o privadas, ya sea en el área financiera, retail, entre otras, por lo que se han desarrollado una multiplicidad de estrategias y acciones para abordarla.

El presente trabajo de título se realiza en el contexto del Área de Desarrollo Empresarial de la Dirección de Compras y Contratación Pública (Dirección ChileCompra). Su objetivo es contribuir al desarrollo de la cartera de proveedores del sistema de Mercado Público, a través del estudio del comportamiento de inactividad temporal (fuga) mediante el análisis de la información transaccional, permitiéndole al Área de Desarrollo Empresarial, contar con un diseño focalizado y eficiente de sus acciones buscando optimizar el match oferta-demanda en el Mercado Público.

La solución propuesta se compone de la construcción de un modelo de comportamiento según estados, del diseño de un criterio para caracterizar el problema de la fuga transaccional en el sistema y finalmente el desarrollo de un modelo de clasificación que permita predecir el comportamiento transaccional de los proveedores en el sistema de Mercado Público.

Para abordar el problema, se trabaja en el ámbito de la Minería de Datos, adoptando la metodología CRISP-DM (Cross Industry Standard Process for Data Mining). Metodología comúnmente empleada en la literatura para este tipo de problemas.

El modelo desarrollado corresponde a una herramienta para la toma de decisiones del Área de Desarrollo Empresarial, pues como se describe en el presente estudio; la información transaccional de la plataforma de Mercado Público es un adecuado predictor de comportamiento. De este modo se pueden potenciar las acciones del área; ya que, al identificar a los proveedores con probabilidad de deserción, es posible retenerlos y cumplir con uno de los grandes objetivos del sistema: garantizar la eficiencia en el Mercado de las Compras Públicas.

Mis agradecimientos:

Con mucha alegría escribo estos agradecimientos. Fueron años de esfuerzo, constancia, valor y coraje.

Primero, lo más importante. Comenzar agradeciendo a mi Familia. A Julio, mi padre, compañero y maestro. Gracias por tu apoyo, ejemplo de vida, consejos y bendita aura de paz. A Carmen, mi madre, abnegada, esforzada y siempre preocupada. Gracias por tus cuidados, amor y entrega generosa. A Vanessa, mi hermana, sensible y multifacética. Gracias por tu delicadeza, cariño, apoyo y paciencia. Por ustedes mi vida, por ustedes todo!.

No puedo dejar de lado a los amigos, aquellos que la vorágine de la vida determina caminos diversos, pero que pese a ello jamás dejaron de alentar, apoyar y acompañar. Esos amigos de verdad que no necesitan tiempo, hora ni lugar para ser.

Por otro lado están aquellos amigos que me acompañaron en San Francisco de los Pajaritos. Ejemplos de lealtad, fe y esperanza. Personas maravillosas que me mostraron lo mejor del ser humano con su compañía.

A los hijos de Bello, compañeros en el camino de la construcción del futuro. A esos ingenieros que buscan hacer de este mundo un lugar más ameno, acogedor y humano. Especialmente al CMS, equipo humano con quienes profesamos la ingeniería del amor y que le han dado un real sentido a lo que es el “ser ingeniero”.

Y finalmente al gran equipo del Área de Desarrollo Empresarial de ChileCompra, que me brindo la oportunidad de realizar este trabajo y de aprender algo muy importante que no aprendí en la universidad; valorar mis capacidades.

Solo me resta dar gracias a Dios por los altos y bajos, las pruebas y bendiciones.

Es hora de nacer de nuevo...

“Los problemas sociales son morales, pero no son sólo morales: encarnan también problemas técnicos que han de ser resueltos para poder aplicar normalmente los principios.

Si los salarios no alcanzan para la vida, la moral enseña que hay que hacerlos tales que alcancen. Pero, ¿por qué medios?, ¿Produciendo una deflación, una inflación, para dar más trabajo, abriendo nuevas industrias, señalando precios a los productos?... Todas esas medidas deben ser estudiadas bajo el punto de vista técnico y de eficacia.

El Evangelio es indispensable, sin él no hay solución; pero jamás enseñó Jesús que quedaban los hombres dispensados de estudiar las soluciones prudentiales, antes al contrario las urgió con rara vehemencia y de ellas nos pedirá cuenta en proporción a la capacidad para descubrirlas.

Parece que es necesario insistir en ese punto, pues es frecuente el pecado de pereza y en todas partes se echa de menos equipos de hombres bien formados en los principios y no menos preparados en la técnica que resuelvan los complicados problemas del mundo en vías de crecimiento”

San Alberto Hurtado, S.J. (Hurtado, 2004)

INDICE DE CONTENIDOS

1. CAPITULO 1: INTRODUCCION	1
1.1. Antecedentes Generales	2
1.2. Justificación del Trabajo	3
1.3. Objetivos	5
1.3.1. Objetivo General	5
1.3.2. Objetivos Específicos	5
1.4. Alcances del estudio	5
1.5. Apoyos Institucionales.....	6
1.6. Metodología	6
1.6.1. Carta Gantt	8
2. CAPITULO 2: DIRECCIÓN CHILECOMPRA.....	9
2.1.1. Estrategia 2010 - 2012.....	10
2.1.2. Objetivos Estratégicos	10
2.1.3. Área de Desarrollo Empresarial	11
2.2. El problema de la “fuga” en el sistema de Mercado Público	11
2.2.1. Criterios para definir fuga.....	15
2.2.2. Diagrama de Estados de un proveedor en el Sistema.....	15
2.2.3. Criterio de fuga: “Semestre Móvil de Actividad” (SMA)	17
2.2.4. Análisis de la fuga en el sistema.....	18
2.2.5. Consecuencias del problema de la fuga en el sistema	20
3. CAPITULO 3: MARCO CONCEPTUAL	23
3.1. Metodología de Minería de datos	24
3.1.1. Metodología CRISP-DM.....	25
3.2. Técnicas de Minería de datos	26
3.2.1. Clustering.....	29
3.2.1.1. Algoritmo K-medias	29
3.2.1.2. Algoritmo EM	30
3.2.2. Clasificación.....	31
3.2.2.1. Árboles de Decisión.....	32

3.2.2.2.	Reglas de Clasificación	32
3.3.	Software de Minería de Datos	33
3.3.1.	WEKA	33
3.3.1.1.	Test y Evaluación en Weka	34
4.	CAPITULO 4: COMPRESION DE LOS DATOS	36
4.1.	Descripción de los Datos:.....	36
4.2.	Set de datos seleccionado	38
4.3.	Limpieza y transformación de la Información	39
4.4.	Exploración de las variables.....	42
5.	CAPITULO 5: MODELADO	49
5.1.	Aplicación de Técnicas de Clusterización:	49
5.1.1.	Empresas Micro	51
5.1.2.	Empresas Pequeñas.....	52
5.1.3.	Empresas Medianas	54
5.1.4.	Empresas Grandes	56
5.1.5.	Comentarios generales sobre la clusterización.....	57
5.1.6.	Gráficos de dispersión de clústeres	58
5.2.	Aplicación de Técnicas de Selección de Atributos	61
5.3.	Aplicación de Técnicas de Clasificación.....	63
5.3.1.	Arboles de Decisión	63
5.3.2.	Reglas de Decisión	66
5.4.	Evaluación de la Clasificación Realizada	66
5.4.1.	Empresas Micro	67
5.4.2.	Empresas Pequeñas.....	67
5.4.3.	Empresas Medianas	68
5.4.4.	Empresas Grandes	68
5.4.5.	Comentarios generales sobre la evaluación	69
6.	CAPITULO 6: EVALUACIÓN Y APLICACIONES	70
6.1.	Aplicación de modelo de clasificación	73
6.2.	Aplicaciones del modelo:.....	74

7. CAPITULO 7: CONCLUSIONES	79
7.1. Conclusiones.....	79
7.2. Mejoras al modelo, trabajo futuro y comentarios generales del sistema.....	81
8. CAPITULO 8: BIBLIOGRAFÍA	84
9. CAPITULO 9: ANEXOS	87
9.1. Anexo A: Tipos de Contratación.....	87
9.2. Anexo B: ChileProveedores	89
9.3. Anexo C: Área de Desarrollo Empresarial.....	90
9.4. Anexo D: Criterio MVI: “Máxima Ventana de Inactividad”	91
9.5. Anexo E: Demostración Intervalo Óptimo para Criterio SMA.	92
9.6. Anexo F: Implementación de Técnicas de Análisis de Datos en Weka.....	93
9.7. Anexo G: Selección de Atributos.....	95
9.8. Anexo H: Arboles de Decisión.....	99
9.9. Anexo I: Reglas de Decisión	100

INDICE DE TABLAS, ILUSTRACIONES y ECUACIONES

Tablas:

TABLA 1: MATRIZ DE FACTORES QUE INCIDEN EN LA FUGA DE PROVEEDORES.....	14
TABLA 2: EVOLUCIÓN DE ESTADOS EN EL TIEMPO (BASE ACTIVOS 2008).....	19
TABLA 3: EVOLUCIÓN DE ESTADOS EN EL TIEMPO, BASE ACTIVOS 2008 (TASA DE FUGA)	20
TABLA 4: INDICADORES DE FIDELIZACIÓN DE CLIENTES.....	21
TABLA 5: LÍNEA DE BASE PARA ANÁLISIS DE DATOS	36
TABLA 6: DATOS PARA EXTRAER DESDE SISTEMAS.....	37
TABLA 7: SET DE VARIABLES	38
TABLA 8: PROVEEDORES QUE SERÁN ELIMINADOS DEL ESTUDIO	40
TABLA 9: PREPROCESADO LÍNEA DE BASE.....	40
TABLA 10: DATASET FINAL Y SUS PORCENTAJES DE CALIDAD DE DATOS.....	41
TABLA 11: VARIABLE: ESTADO PROVEEDORES	42
TABLA 12: VARIABLE: TAMAÑO EMPRESA	42
TABLA 13: VARIABLE: ANTIGÜEDAD EN TRANSACCIÓN.....	43
TABLA 14: VARIABLE: CANTIDAD DE ADJUDICACIONES	43
TABLA 15: VARIABLE: CANTIDAD DE ÓRDENES DE COMPRA.....	44
TABLA 16: VARIABLE: CANTIDAD DE OFERTAS	44
TABLA 17: VARIABLE: MONTO TRANSADO	45
TABLA 18: VARIABLE: CANTIDAD DE OFERTAS RECHAZADAS.....	45
TABLA 19: VARIABLE: PROMEDIO LOGEO	46
TABLA 20: VARIABLE: REGISTRO EN CHILEPROVEEDORES	46
TABLA 21: VARIABLE: VIGENTE EN CHILEPROVEEDORES	46
TABLA 22: VARIABLE: RECLAMOS.....	47
TABLA 23: VARIABLE: CANTIDAD DE MESES LOGEADO	47
TABLA 24: VARIABLE: TASA OF-RATE	48
TABLA 25: VARIABLE: TASA WIN-RATE	48
TABLA 26: SEGMENTACIÓN TAMAÑO EMPRESA SEGÚN VENTAS ANUALES.....	49
TABLA 27: NÚMERO DE CLUSTERS POR SEGMENTO EMPRESA	50
TABLA 28: CLUSTERS: EMPRESAS MICRO.....	51
TABLA 29: CLUSTERS: EMPRESAS PEQUEÑAS	52
TABLA 30: CLUSTERS: EMPRESAS MEDIANAS	54
TABLA 31: CLUSTERS: EMPRESAS GRANDES	56
TABLA 32: RESUMEN CLUSTERS	58
TABLA 33: VARIABLES SELECCIONADAS PARA CLASIFICACIÓN	62
TABLA 34: CLUSTER PROBABLE Y DISTRIBUCIÓN DE PROBABILIDADES	65

TABLA 35: CANTIDAD DE REGLAS POR TAMAÑO DE EMPRESA.....	66
TABLA 36: MATRIZ DE CONFUSIÓN MICRO	67
TABLA 37: MATRIZ DE CONFUSIÓN PEQUEÑA.....	67
TABLA 38: MATRIZ DE CONFUSIÓN MEDIANA.....	68
TABLA 39: MATRIZ DE CONFUSIÓN GRANDE.....	68
TABLA 40: EVALUACIÓN AL AÑO 2010 (BASE 2009) (EMPRESAS MICRO).....	70
TABLA 41: EVALUACIÓN AL AÑO 2010 (BASE 2009) (EMPRESAS PEQUEÑA).....	71
TABLA 42: EVALUACIÓN AL AÑO 2010 (BASE 2009) (EMPRESAS MEDIANA).....	72
TABLA 43: EVALUACIÓN AL AÑO 2010 (BASE 2009) (EMPRESAS GRANDES).....	73
TABLA 44: EVALUACIÓN PROVEEDORES NUEVOS 2010.....	74
TABLA 45: PROVEEDORES BASE 2009 TRANSANDO A JUNIO 2010	75
TABLA 46: OPCIONES DE CONFIGURACIÓN K-MEANS EN WEKA.....	93
TABLA 47: OPCIONES DE CONFIGURACIÓN EM EN WEKA	93
TABLA 48: OPCIONES DE CONFIGURACIÓN ALGORITMO J48	93
TABLA 49: OPCIONES DE CONFIGURACIÓN ALGORITMO PART.....	94
TABLA 50: SELECCIÓN DE ATRIBUTOS EMPRESAS MICRO.....	95
TABLA 51: SELECCIÓN DE ATRIBUTOS EMPRESAS PEQUEÑAS	96
TABLA 52: SELECCIÓN DE ATRIBUTOS EMPRESAS MEDIANAS	97
TABLA 53: SELECCIÓN DE ATRIBUTOS EMPRESAS GRANDES	98

Ilustraciones:

ILUSTRACIÓN 1: PROVEEDORES INSCRITOS EN CHILECOMPRA ACUMULADO 2003 A 2009	2
ILUSTRACIÓN 2: PROVEEDORES ACTIVOS EN CHILECOMPRA AÑOS 2003 A 2009	3
ILUSTRACIÓN 3: TASA DE CRECIMIENTO DE PROVEEDORES INSCRITOS EN MERCADO PÚBLICO	4
ILUSTRACIÓN 4: CARTA GANTT DEL PROYECTO	8
ILUSTRACIÓN 5: MEDIDAS DE SATISFACCIÓN POR ESTADO.....	13
ILUSTRACIÓN 6: ÁRBOL DE HIPÓTESIS DEL PROBLEMA DE LA FUGA	13
ILUSTRACIÓN 7: GRAFO DE ESTADOS DE UN PROVEEDOR EN EL SISTEMA.....	17
ILUSTRACIÓN 8: CRITERIO DE FUGA “SEMESTRE MÓVIL DE ACTIVIDAD” (SMA).....	17
ILUSTRACIÓN 9: PROVEEDORES EN ESTADO (BASE ACTIVOS 2008).....	18
ILUSTRACIÓN 10: PROVEEDORES EN ESTADO. BASE ACTIVOS 2009.....	19
ILUSTRACIÓN 11: PORCENTAJE DE PROVEEDORES EN ESTADOS EN EL TIEMPO	20
ILUSTRACIÓN 12: FASES DE UN PROCESO KDD	23
ILUSTRACIÓN 13: ENCUESTA SOBRE METODOLOGÍAS EMPLEADAS EN PROYECTOS DE DATA MINING.....	25
ILUSTRACIÓN 14: FASES DEL PROCESO CRISP-DM	25
ILUSTRACIÓN 15: TÉCNICAS UTILIZADAS PARA REALIZAR MINERÍA DE DATOS (2007).....	27

ILUSTRACIÓN 16: TÉCNICAS DE MINERÍA DE DATOS	27
ILUSTRACIÓN 17: PROCESO DE CLUSTERING CON DOS ATRIBUTOS.	30
ILUSTRACIÓN 18: VARIABLE: ESTADO PROVEEDORES	42
ILUSTRACIÓN 19: VARIABLE: TAMAÑO EMPRESA	42
ILUSTRACIÓN 20: VARIABLE: ANTIGÜEDAD EN TRANSACCIÓN	43
ILUSTRACIÓN 21: VARIABLE: CANTIDAD DE ADJUDICACIONES.....	43
ILUSTRACIÓN 22: VARIABLE: CANTIDAD DE ÓRDENES DE COMPRA	44
ILUSTRACIÓN 23: VARIABLE: CANTIDAD DE OFERTAS	44
ILUSTRACIÓN 24: VARIABLE: MONTO TRANSADO.....	45
ILUSTRACIÓN 25: VARIABLE: CANTIDAD DE OFERTAS RECHAZADAS	45
ILUSTRACIÓN 26: VARIABLE: PROMEDIO LOGEO.....	46
ILUSTRACIÓN 27: VARIABLE: REGISTRO EN CHILEPROVEEDORES	46
ILUSTRACIÓN 28: VARIABLE: VIGENTE EN CHILEPROVEEDORES.....	46
ILUSTRACIÓN 29: VARIABLE: RECLAMOS	47
ILUSTRACIÓN 30: VARIABLE: CANTIDAD DE MESES LOGEADO	47
ILUSTRACIÓN 31: CLUSTER V/S WIN-RATE (MICROEMPRESAS).....	59
ILUSTRACIÓN 32: CANTIDAD DE ADJUDICACIONES V/S WINRATE (GRANDES EMPRESAS)	60
ILUSTRACIÓN 33: ARBOLES DE DECISIÓN POR TAMAÑO DE EMPRESA.....	63
ILUSTRACIÓN 34: EJEMPLO DE DATOS NO CLASIFICADOS	64
ILUSTRACIÓN 35: CRITERIO DE FUGA “MÁXIMA VENTANA DE INACTIVIDAD” (MVI)	91
ILUSTRACIÓN 36: MÁXIMA VENTANA INACTIVA EN PERIODO ACTIVO	92
ILUSTRACIÓN 37: CANTIDAD DE VECES QUE UN PROVEEDOR POSEE UN CICLO	92
ILUSTRACIÓN 38: EJEMPLO DE RAMA DE ÁRBOL DE CLASIFICACIÓN.....	99

Ecuaciones:

ECUACIÓN 1: TASA DE DESERCIÓN	21
ECUACIÓN 2: TASA DE RETENCIÓN.....	21
ECUACIÓN 3: VIDA MEDIA DE UN CLIENTE CON LA EMPRESA.....	21
ECUACIÓN 4: MEDIDA DE DISTANCIA EUCLIDEANA	29
ECUACIÓN 5: CÁLCULO DE VEROSIMILITUD EN ALGORITMO EM.....	31
ECUACIÓN 6: FORMULA DE OBTENCIÓN DE F-MEASURE	35

1. CAPITULO 1: INTRODUCCION

La Dirección de Compras y Contratación Pública (Dirección ChileCompra, 2010) a través de su política de creación de valor en el Mercado Público ha decidido utilizar herramientas de inteligencia de negocios en beneficio de los proveedores del Estado.

Es aquí donde surge el presente trabajo de título, que busca generar conocimiento acerca del comportamiento de los Proveedores del sistema mediante el uso de información transaccional de la plataforma de Mercado Público. En particular, estudiar lo referido al comportamiento de prolongada inactividad, lo que se denominará como fuga.

Actualmente, los indicadores de inactividad en el sistema rondan el 40% a los 6 meses y el 50% al año desde la última transacción. Con la finalidad de disminuir este porcentaje y poder garantizar al sistema que el Mercado Público se consolide como un mercado transparente, eficiente y equitativo, y que cumpla con los objetivos estratégicos de incrementar la efectividad del encuentro entre la oferta y la demanda, se busca detectar qué tipo de proveedores son los más propensos a “fugarse” del sistema y, con ello contribuir al desarrollo eficiente de la cartera de proveedores del Estado.

Al contar con una adecuada descripción de los diversos tipos de proveedores existentes en la actualidad (no sólo de los fugados), la Dirección ChileCompra podrá tomar acciones concretas para invitar a participar a cada grupo y así cumplir con los estándares de eficiencia y sustentabilidad del mercado, pues el grado de participación de proveedores en las licitaciones del estado es la variable más aceptada a nivel internacional para medir la salud del sistema de contratación pública (Dirección ChileCompra, 2008).

Para abordar el problema se utilizarán como principales fuentes de información, los registros almacenados en la base de datos de la Dirección ChileCompra y el DataWarehouse operacional desde el año 2007 a 2009, buscando una adecuada descripción del comportamiento de los proveedores del Estado. Esto se realizará a través del empleo de técnicas de Minería de Datos siguiendo la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) (CRISP-DM, 2010).

Para efectos de la construcción del modelo, primero se definirá el conjunto de variables adecuado para el estudio; luego se procederá a realizar el procesamiento de la información; etapa que se estima consume la mayor cantidad del tiempo del estudio.

Finalmente, se aplicarán técnicas especializadas de Minería de Datos con el fin de obtener información relevante, la cual será interpretada, con miras al diseño de una estrategia de gestión de cartera de proveedores.

1.1. Antecedentes Generales

El sistema de Mercado Público (Dirección ChileCompra, 2010) es administrado por la Dirección de Compras y Contratación Pública. Es un sistema de información público y gratuito, tal como lo indica el artículo 19 de la Ley de Compras y Contratación Pública

“Artículo 19.- Créase un Sistema de Información de Compras y Contrataciones de la Administración, a cargo de la Dirección de Compras y Contratación Pública, que se aplicará a los organismos señalados en el artículo 1º de la presente ley, y que deberá estar disponible a todo el público, en la forma que regule el reglamento. El Sistema de Información será de acceso público y gratuito”. (Ministerio de Hacienda, 2003)

La naturaleza de “público y gratuito” del sistema, implica que no existen barreras de entrada en el registro de proveedores en la plataforma de Mercado Público. Desde 2003 a la fecha ya son más de 350.000 las empresas que se han registrado.

Ilustración 1: Proveedores Inscritos en ChileCompra Acumulado 2003 a 2009



(Fuente: Balance Anual (Dirección ChileCompra, 2008))

Dado esto, el comportamiento relevante a observar para el presente estudio no es el registro en la plataforma, sino que la actividad en ella. Según datos de 2009 cerca de 90.000 empresas registran anualmente transacciones.

Definiremos una **transacción** en el Mercado Público al envío de una oferta¹ o la adjudicación de una orden de compra². Se entenderá entonces como **proveedor**

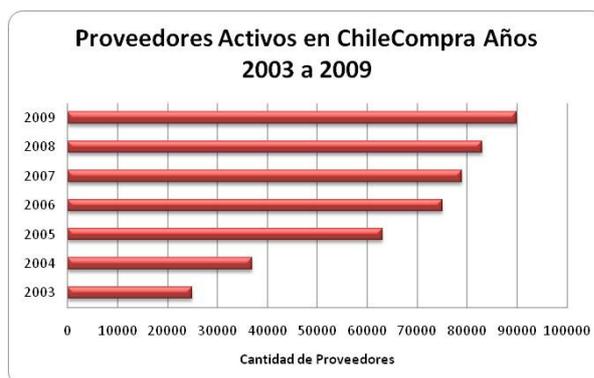
¹ Nota: Oferta en el sistema se refiere al proceso en el cual el proveedor del estado decide realizar una propuesta para satisfacer la necesidad de una adquisición publicada en el sistema de Mercado Público. (Dirección ChileCompra, 2010)

² Nota: La Orden de Compra es el documento mediante el cual se formaliza la adquisición con el proveedor cuya oferta ha sido aceptada y adjudicada. (Dirección ChileCompra, 2010)

activo, a aquella empresa que ha realizado por lo menos una transacción en el sistema en un periodo de tiempo determinado.

Es importante destacar que existen diversos mecanismos de adjudicación, en algunos de ellos no es necesario el envío de ofertas para adjudicar. Detalle de estos mecanismos se encuentra en el Anexo A

Ilustración 2: Proveedores Activos en ChileCompra Años 2003 a 2009



(Fuente: Balance Anual (Dirección ChileCompra, 2008))

La diferencia entre el número de empresas proveedoras registradas en la plataforma y el número de empresas transando en el Mercado Público está determinada por el comportamiento de los proveedores, principalmente por la inactividad (fuga), la cual será el principal objeto de análisis a lo largo del estudio.

1.2. Justificación del Trabajo

En el marco estratégico planteado por la Dirección ChileCompra al 2012, surge la necesidad de estudiar el comportamiento de proveedores como una consideración básica para la toma de decisiones. Se explicita en los desafíos a 2012 la necesidad de “conocer más a los clientes proveedores³, para ofrecerles soluciones específicas a sus necesidades y tener más Mercado Público para todos” (Dirección ChileCompra, 2009).

Se ha definido que la generación de conocimiento apunta a conocer más a los proveedores, descubrir los patrones que los caracterizan en el sistema, asumiendo que la información transaccional es un buen referente para segmentar, tal como se enuncia en el *paper* “Caracterización del proceso de fuga de clientes utilizando información transaccional” (Segovia, y otros, 2005).

³ Se hace la distinción de “cliente proveedor” ya que en rigor la Dirección ChileCompra posee dos tipos de clientes. Las Empresas proveedoras del Estado y los Servicios Públicos que utilizan el sistema.

Existen múltiples comportamientos al interior del sistema. Uno de ellos en particular es la fuga o mejor dicho inactividad en el sistema. La diferencia entre ambos conceptos radica en que no es posible hablar de fuga entendido como deserción definitiva del sistema

Así como no existen barreras de entrada, tampoco existen barreras de salida. Por ende, solo es posible definir periodos de inactividad temporal para la historia de cada proveedor (los cuales pueden prolongarse en el tiempo indefinidamente). Surge así entonces, la necesidad de diseñar un criterio de fuga adecuado para el análisis.

La inactividad en promedio bordea el 40 al 45% de los proveedores a los 6 meses desde su última transacción. Dado que se requiere mantener al sistema con una oferta constante de proveedores satisfaciendo las demandas de los organismos públicos, la deserción de usuarios proveedores representa un problema para la institución, quien para garantizar la eficiencia del mercado define tomar acciones concretas para minimizar este efecto.

La cartera de proveedores es muy amplia. Por ende se vuelve estrictamente necesaria la elaboración de estrategias focalizadas por segmentos, con la finalidad de abordar puntos críticos de la oferta de proveedores logrando maximizar el impacto que el Área de Desarrollo Empresarial pueda generar con sus acciones.

Se pretende abordar el problema en particular con técnicas de minería de datos; replicando estudios similares en otros ámbitos, como lo son el sector bancario, seguros, salud, población criminal, etc. Se describirá esta bibliografía en el capítulo 3 al analizar los alcances de la minería de datos.

Finalmente, la tasa de crecimiento de la inscripción de proveedores en el sistema disminuye en el tiempo (Gráfico 3), lo que incrementa la relevancia del trabajo focalizado con los clientes-usuarios actuales del sistema, es decir, fidelización de la cartera de proveedores.

Ilustración 3: Tasa de Crecimiento de Proveedores inscritos en Mercado Público



(Fuente: Elaboración Propia en base a datos entregados por la institución)

1.3. Objetivos

1.3.1. Objetivo General

Para el estudio a realizar se ha determinado el siguiente objetivo general:

“Diseño de un modelo de predicción de comportamiento de inactividad temporal (fuga), mediante el análisis de la información transaccional, para contar con un diseño de acciones focalizadas y contribuir al eficiente *match* entre la oferta y la demanda”.

1.3.2. Objetivos Específicos

- Construir un modelo de comportamiento según estados que permita clasificar a los proveedores en función de su actividad transaccional en la plataforma.
- Identificar qué tipo de proveedores son los más propensos a la permanencia en el sistema así como a los proveedores que tienden a la fuga.
- Entregar un conjunto de predictores de comportamiento de proveedores al Área de Desarrollo Empresarial de ChileCompra, que permita la confección de planes de acción adecuados en función a las estrategias de la organización.
- Diseñar propuestas de acción focalizadas por segmentos, donde sea posible.

1.4. Alcances del estudio

Dado que no existen antecedentes previos en lo referido a la generación de conocimiento del comportamiento de proveedores en el sistema de Mercado Público, ni una caracterización de estados, más allá de indicadores de gestión como cantidad de proveedores transando, ofertando, número de ofertas enviadas, órdenes de compra generadas, etc. se pretende generar una nueva clasificación del estado de un proveedor en el Mercado Público, utilizando reglas de negocio, de modo que un proveedor registrado, pueda ser clasificado dentro de una categoría específica durante su permanencia en el sistema.

Además, se busca encontrar patrones de comportamiento recurrentes entre los proveedores para así diseñar una primera versión de un modelo que prediga el comportamiento de los proveedores entre estados.

Dado que, la participación de los proveedores en el sistema es voluntaria, libre, sin barreras de entrada (salvo para el registro en el sistema de ChileProveedores)⁴, el

⁴ Existen barreras para la inscripción en el Registro de Proveedores del Estado (ChileProveedores, 2010),

problema a investigar debe considerar que los proveedores pueden o no estar participando en otros mercados y que en rigor la fuga no existe, sino que se trata de una “inactividad temporal”. La **fuga** surge como un criterio de negocio según la necesidad estratégica de la institución.

Entonces, se busca determinar si es posible, a través de la información transaccional de la plataforma, establecer alertas sobre los cambios de estado de los proveedores en el mismo sistema. En otras palabras se busca validar la hipótesis de que “el comportamiento transaccional de los proveedores es un buen predictor de comportamiento en el Mercado Público”.

El horizonte de tiempo del estudio contempla el uso de datos históricos comprendidos principalmente en el año 2009, para finalmente caracterizar el comportamiento de proveedores en el año 2010.

Es posible que el estudio al contar con un intervalo de tiempo acotado omita información histórica de proveedores (que participan desde la creación del sistema en 2003). Esta información no se considera pues no existe una base de datos con la confiabilidad necesaria. Se adopta el supuesto de que esta información no es relevante para la caracterización.

1.5. Apoyos Institucionales

Para el desarrollo de este estudio se cuenta con los apoyos institucionales necesarios. El Área de Desarrollo Empresarial a través de su Jefatura y equipo de trabajo han contribuido no solo con la información necesaria, sino que además han manifestado apoyo en lo referido a la validación de las interpretaciones necesarias, especialmente en la etapa de evaluación del modelo. Por otro lado, se han facilitado los accesos necesarios a las bases de datos operacionales para la extracción de la información necesaria.

Se cuenta además con el apoyo para involucrar en las líneas de trabajo del Área las iniciativas y modelos que el presente trabajo de título manifiesta. Esto se materializará a través de la implementación de las propuestas en el plan de trabajo del Área durante el segundo semestre de 2010.

1.6. Metodología

Las actividades a realizar se enmarcan dentro de una metodología predefinida para abordar este tipo de problemas. La metodología seleccionada corresponde a CRISP-DM (CRISP-DM, 2010), la cual se definirá con mayor detalle en el capítulo 3.

como por ejemplo las inhabilidades para transar con el estado. Detalle en Anexo B

A continuación se enuncia una descripción de las principales actividades a realizar y posteriormente su distribución en el tiempo en el formato de Carta Gantt⁵.

Etapa 1: Análisis del problema: Analizar el comportamiento transaccional de proveedores a través del diseño de un modelo que los caracteriza según estados. Estudiar el Problema de la Fuga: Describir el problema, caracterizar los niveles de fuga en el sistema y definir un criterio de negocio válido para referirse al fenómeno. Se contempla un trabajo intenso con las áreas de negocio relacionadas en la institución. Ambos análisis se describirán en el Capítulo 2.

Por otro lado, al definirse la aplicación de Minería de Datos como la técnica para el análisis, se contempla la documentación bibliográfica a nivel conceptual relevante y la búsqueda de la herramienta de software necesaria para el análisis. Además se pretende agendar reuniones con expertos en la materia con la finalidad de validar avances. En el capítulo 3 se describirá el marco teórico de la investigación.

Etapa 2: Comprensión de los datos: Definición de Variables a observar: investigación de las bases de datos, data warehouses y fuentes de datos relevantes para el estudio.

Etapa 3: Preparación de los datos: Extracción de los Datos: Consolidar la información, extraer lo necesario y construir una base de datos con las variables anteriormente definidas. Limpieza de la Información: Corrección de los datos, eliminación de outliers, valores faltantes, etc.

Exploración de los datos: Describir los datos, sus distribuciones, y en general, su composición. Selección de la muestra de datos: Definir un criterio de selección de muestra para realizar el modelado a través de un trabajo conjunto con el Área de Desarrollo Empresarial.

Estas subetapas, junto a la comprensión de los datos, serán descritas en el capítulo 4 de este informe.

Etapa 4: Modelado

Aplicación de Técnicas de Clusterización: Los algoritmos de Clusterización buscan grupos de instancias con características similares, según algún criterio de comparación de valores definido por el algoritmo. En este caso la clusterización se empleará para definir conglomerados de proveedores con características comunes.

Aplicación de Técnicas de Selección de Atributos: Se escogerán algunas de las variables con la finalidad de eliminar redundancias (por ejemplo correlaciones, datos que no aporten al modelo, etc.). No solo se tomará en cuenta el resultado de los

⁵ http://es.wikipedia.org/wiki/Diagrama_de_Gantt

algoritmos sino que además se discutirá con miembros del área la pertinencia de la agregación o eliminación de variables del modelo.

Aplicación de Técnicas de Clasificación: La clasificación es en la práctica de la minería de datos el problema de mayor frecuencia. Se construirá un modelo que permita predecir la categoría de las instancias, en función de los datos de entrada. Esta clasificación permitirá obtener la probabilidad de caer en cierta categoría a una instancia evaluada.

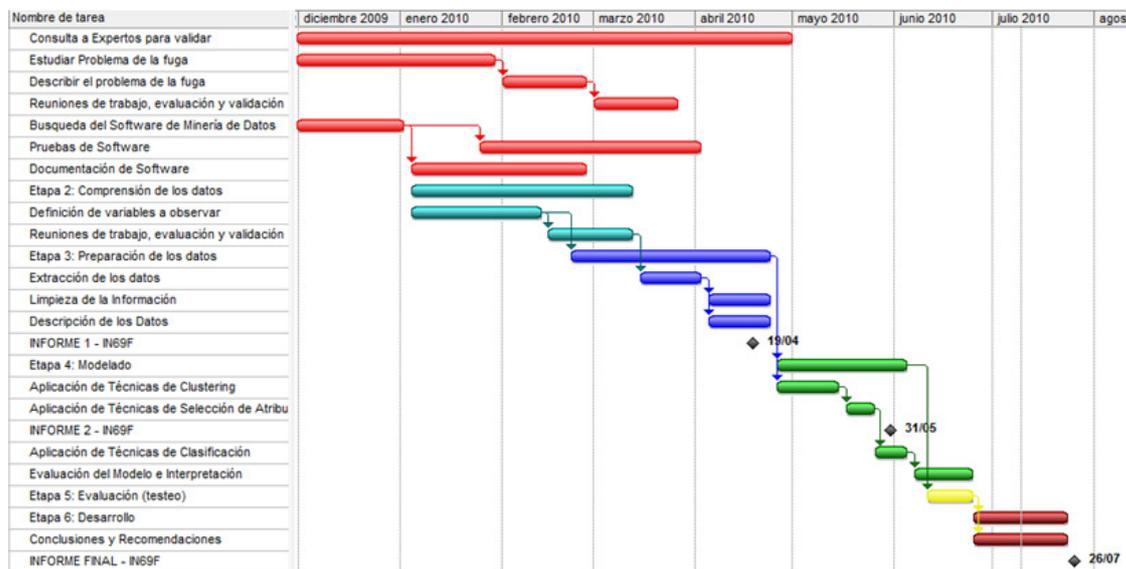
Se realizará inducción sobre el atributo “cluster” obtenido anteriormente, generando reglas de decisión sobre los atributos. El capítulo 5 abordará esta etapa.

Etapa 5: Evaluación: Principalmente esta etapa tiene que ver con la evaluación de los modelos anteriormente diseñados, en cuanto a calidad y certeza de caracterización. Se evaluará a través de la eficacia de los algoritmos, con expertos y equipo del área de Desarrollo Empresarial. Esta etapa también se incluirá en el capítulo 5.

Etapa 6: Desarrollo: Realizar una interpretación objetiva y fundada en los datos obtenidos del comportamiento de proveedores, junto con el diseño de una propuesta de acción focalizada por segmento. Capítulo 6.

1.6.1. Carta Gantt

Ilustración 4: Carta Gantt del Proyecto



(Fuente: Elaboración Propia)⁶

⁶ Confeccionado en Software Microsoft Office Project Standard 2007 (trial version). <http://www.microsoft.com/>

2. CAPITULO 2: DIRECCIÓN CHILECOMPRA

La Dirección de Compras y Contratación Pública (Dirección ChileCompra) planifica, desarrolla y promueve las iniciativas necesarias para que el Mercado Público crezca armónicamente. Es un Servicio Público descentralizado, dependiente del Ministerio de Hacienda, y sometido a la supervigilancia del Presidente de la República. Se creó con la Ley de Compras Públicas (Ministerio de Hacienda, 2003) y comenzó a operar formalmente el 29 de agosto de 2003, como sucesor de la D.A.E (Dirección de Aprovisionamiento del Estado).

En el mercado de las compras públicas se produce la interacción de las necesidades por bienes y servicios de los organismos estatales con la diversidad de oferta de miles de proveedores interesados en detectar oportunidades de negocio, generando transacciones que en 2009 superaron los 6.000 millones de dólares (equivalente a un 3% de Producto Interno Bruto del país).

Durante sus seis años de existencia, la Dirección ChileCompra ha logrado ordenar y sistematizar un sistema esencialmente complejo, donde confluyen los infinitos intereses, demandas y necesidades de los actores que dan vida al Mercado Público, mediante el establecimiento de reglas que ayuden a mejorar la competitividad, aumentando la calidad y transparencia de la información.

La Dirección ChileCompra ofrece una serie de servicios y plataformas electrónicas. A través de su sitio web www.mercadopublico.cl, la plataforma de comercio electrónico más grande del país. Al 2009 transan cerca de 850 organismos del Estado y participan activamente más de 90 mil empresas, generando ahorros para el estado por cerca de US\$ 180 millones y permitiendo una mayor transparencia, eficiencia y acceso a este mercado (Dirección ChileCompra, 2009).

Actualmente están disponibles las tiendas ChileCompra Express, que durante 2009 ha transado US\$929 millones, siendo la mayor tienda en línea del país- y ChileCompra Libros (orientada al sector educación principalmente) que a la fecha ofrece 40 mil títulos y ha demostrado ser un canal eficaz para realizar adquisiciones en el ámbito de la cultura.

Entre los servicios disponibles se cuenta la operación de 30 Centros de Emprendimiento ChileCompra ubicados en las principales ciudades del país. Asimismo, en www.chilecompra.cl, pueden acceder a soporte, formación, material de apoyo y completos reportes de transacciones para analizar las oportunidades de negocios de la plataforma de licitaciones www.mercadopublico.cl.

La Dirección ChileCompra también administra el Registro Electrónico de Proveedores del Estado (ChileProveedores, 2010), en el cual 40.000 empresas ya están inscritas y que se abrió en agosto de 2009 al sector privado.

2.1.1. Estrategia 2010 - 2012

A continuación se explicita la visión de futuro de la organización y la misión institucional, ambas definidas en el Plan Estratégico 2010 – 2012 (Dirección ChileCompra, 2009) vigente actualmente en la institución.

Visión de Futuro:

“Durante la próxima década la Dirección ChileCompra será una institución reconocida como líder a nivel nacional e internacional en la implementación de un mercado público transparente, participativo y eficiente, generando un gran avance en la modernización de la gestión pública y facilitando la implementación de las políticas públicas. Para esto ofrecerá servicios de calidad, de fácil uso e integrados con los procesos sustanciales de sus clientes”.

Misión Institucional:

“Nuestra misión es crear valor en el Mercado Público, para que los compradores del Estado puedan hacer su trabajo cada día de manera más transparente y eficiente y que los proveedores tengan más facilidad de acceso. Nuestro trabajo está enfocado en nuestros clientes, asegurando una alta calidad, y usando de manera intensiva las tecnologías de información”.

2.1.2. Objetivos Estratégicos

Para el período 2010 – 2012, la Dirección ChileCompra ha definido los siguientes objetivos estratégicos:

Incrementar la efectividad del encuentro entre la oferta y la demanda en el mercado de las compras públicas: El eje del trabajo estará puesto en tratar la información del mercado público como un activo valioso para compradores, proveedores y otros grupos de interés. Esto nos permitirá incrementar el ahorro y llegar a más proveedores ampliando la base de empresas que participan de este mercado, con especial énfasis en aquellas áreas donde la demanda del Estado encuentra una respuesta más débil⁷.

⁷ Se considerará débil un sector demandado cuando existan licitaciones que la Dirección considere “no exitosas”. El criterio para considerar una licitación como no exitosa es que haya sido adjudicada pero con un número de ofertas de proveedores menor a 3. Es un criterio diseñado por la institución que obedece a estándares internacionales.

Incrementar la satisfacción de los clientes del Mercado Público: La estrategia se basa en reducir los costos de transacción entre los compradores y el mercado. Esto implica reconocer las diferencias entre nuestros clientes y de esta forma lograr una mejor segmentación. A partir de esto, se ofrecerán servicios que respondan a las necesidades de cada uno de estos segmentos. Una definición básica es que las soluciones que ofrezca ChileCompra deben ser interoperables⁸.

Instalar los conceptos básicos para un mercado público sustentable y de amplio acceso: Para el desarrollo de este concepto, nos apoyaremos fuertemente en el uso de convenios marco. De manera paulatina se incorporarán criterios de sustentabilidad en las licitaciones de este tipo y de forma complementaria se empezará a mostrar información que permita al comprador orientar su decisión hacia productos y servicios más sustentables. De manera paralela se incorporarán estos conceptos al trabajo con los compradores mediante distintos instrumentos.

El presente trabajo de título apunta directamente en el sentido del primer y segundo objetivo estratégico: incrementar la efectividad del encuentro entre la oferta y la demanda en el mercado de las compras públicas a través de la generación de conocimiento de los clientes y mejorar la satisfacción de ellos a través de políticas dirigidas enfocadas en sus propias necesidades.

2.1.3. Área de Desarrollo Empresarial

La Dirección ChileCompra posee una estructura horizontal en tres niveles. El presente estudio se enmarca dentro del Área de Desarrollo Empresarial, la cual administra directamente la cartera de proveedores registrados en la plataforma de Mercado Público. Esta cartera asciende a más de 350.000 proveedores al 2009.

El Área cuenta con diversos servicios, entre los que se destacan la Red Territorial de Centros de Emprendimiento, el Registro Nacional de Proveedores del Estado (ChileProveedores) y diversos servicios empresariales, entre los que se puede mencionar el confirming y ordering. En el anexo C se detallan estas iniciativas.

2.2. El problema de la “fuga” en el sistema de Mercado Público

La deserción de proveedores de la actividad transaccional o fuga de proveedores se entiende como la acción correspondiente a dejar de participar activamente en el sistema de Mercado Público durante un determinado intervalo de tiempo consecutivo.

⁸ Nota. **Interoperabilidad:** Se espera que la evolución de la Dirección ChileCompra contemple criterios de interoperabilidad natural con los principales sistemas del Estado, incluyendo ERPs, SIGFE, etc. (Dirección ChileCompra, 2009)

Este criterio busca explicitar la distinción entre inactividad (no operatividad transaccional) y fuga, ya que no se puede determinar el momento en que se fuga un proveedor (sistema de libre acceso, sin barreras de entrada ni de salida). A diferencia de instituciones financieras en donde existe un cierre contractual explícito (por ejemplo el cierre de una cuenta corriente), aquí no se puede determinar ese evento. Entonces surge la idea de categorizar según un criterio de negocio.

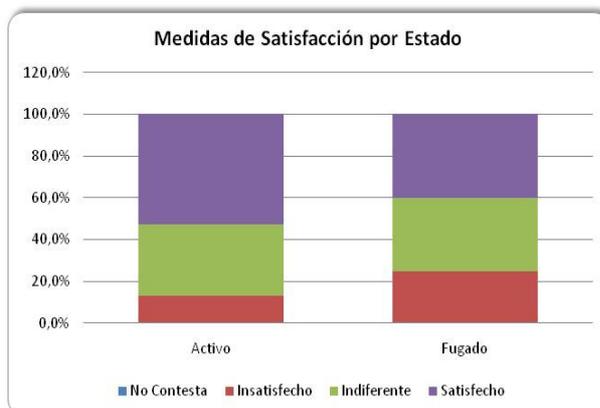
En el ámbito estratégico de la Dirección, se pueden distinguir 3 impactos directos de este problema. En primer lugar, la deserción de proveedores fomenta el incremento de la cantidad de licitaciones no exitosas⁹. Esto se debe a que si no existe la suficiente cantidad de oferentes transando en el sistema, no se puede garantizar el objetivo institucional de eficiencia del mercado de las compras públicas. Por otro lado, es evidente que altos índices de fuga disminuyen la cantidad de proveedores transando lo que fomentaría la concentración de mercados, atentando contra la equidad y la transparencia.

Y por último, existe una relación muy estrecha entre la fuga de proveedores y la satisfacción de ellos con el sistema. Se estima que el fenómeno de la fuga es gatillado por una insatisfacción. Así lo demuestra un análisis realizado con encuestas de satisfacción en la institución, en donde se pudo constatar que la satisfacción neta¹⁰ entre proveedores activos y fugados difiere en 25 puntos porcentuales (los proveedores activos poseen una Satisfacción neta de un 39,8%, contra los fugados que poseen un 15%, como se puede observar en la siguiente gráfica.

⁹ **Licitación no Exitosa:** Se considerará un proceso licitatorio no exitoso desde el punto de vista de la eficiencia al proceso que en su periodo de recepción de ofertas no alcanza el mínimo esperado de 3 ofertas y es adjudicado satisfactoriamente. Es una distinción adoptada por la Dirección ChileCompra basada en estándares internacionales. (Milgrom, 2004)

¹⁰ **Satisfacción neta:** indicador de satisfacción utilizado por la Dirección ChileCompra que consta de la diferencia entre el volumen de proveedores que evalúa positivamente el sistema contra quienes lo evalúan de manera negativa.

Ilustración 5: Medidas de Satisfacción por estado¹¹

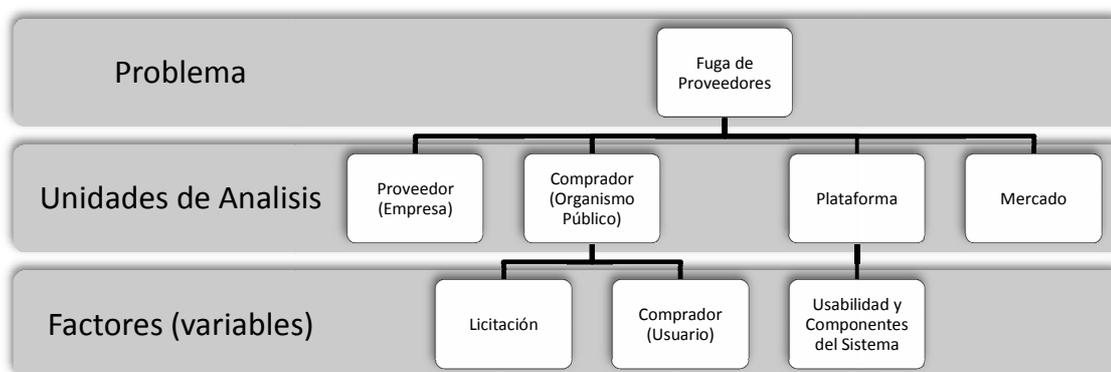


(Fuente: Elaboración Propia en base a información entregada por el Área de Desarrollo Empresarial)

Para analizar la causa del problema de la fuga de proveedores, se ha diseñado un árbol de hipótesis que muestra cuales son los principales factores que eventualmente inciden en la decisión de las empresas de no transar más en el sistema de Mercado Público¹².

Este esquema con hipótesis surge para describir inicialmente el problema de la fuga y se basa en juicio experto consensuado con profesionales del Área.

Ilustración 6: Árbol de Hipótesis del problema de la fuga



(Fuente: Elaboración Propia)

¹¹ Gráfica obtenida de un estudio realizado en el Área de Desarrollo Empresarial a las encuestas de satisfacción de proveedores, cruzando la evaluación del sistema con el nivel transaccional de los usuarios que las responden, buscando perfilar a quienes responden y comprender desde una perspectiva de sistema su evaluación

¹² Algunos de estos factores está demostrado que inciden en el problema. No se desarrollará en profundidad este tema pues no está contemplado dentro de los alcances de la investigación.

Cada una de las unidades de análisis descritas contempla una serie de tópicos que eventualmente pueden ser considerados causales de fuga. Brevemente se describirán algunos de ellos.

Es importante destacar que pueden existir otras variables como por ejemplo la capacidad de las empresas, en términos de volumen de ventas y efectos coyunturales, los cuales son omitidos por simplicidad.

Tabla 1: Matriz de Factores que inciden en la fuga de proveedores

Unidad de Análisis	Factores	Causas
Comprador (Organismo Público)	Licitaciones	Exigencia de boletas de garantía
		Exigencia de marcas específicas (práctica anticompetitiva)
		NO publicación de las ofertas técnicas en licitaciones
		Publicación de Oportunidades de Negocios en plazos no aptos para el normal proceso de oferta o en tiempos que dificultan el diseño de propuestas.
	Usuario Comprador	Prácticas de Compradores (Como la conducta de pago de servicios públicos).
		Modificación de los Tiempos de publicación de licitaciones (Procesos de licitación en estado publicada, sin documentos adjuntos y que no cumplan con fechas óptimas).
		NO publicación de los cuadros finales (Acta de adjudicación).
		Falta de transparencia y probidad (filtrado de licitaciones, concursos dirigidos, etc.).
Mercado	Disminución de Niveles de Oportunidades de negocios.	
	Exigencia de ChileProveedores en procesos de licitación.	
	Proliferación de Convenios Marco.	
Plataforma	Problemas de acceso (navegadores, estabilidad, etc.).	
	Disconformidad con el sistema de búsquedas de Oportunidades de Negocios.	
	Disconformidad con el sistema de inscripción de rubros para recibir información (rubros ChileCompra no calzan con giro de empresas).	
	Generación de Spam (debido al envío de correos por notificaciones masivo).	
	Usabilidad y Estabilidad de la plataforma.	
Proveedores	Problemas de Financiamiento	
	Desinterés y desconfianza del sistema	
	Malas experiencias en el pasado.	

(Fuente: Elaboración Propia)

2.2.1. Criterios para definir fuga

Ya se ha definido la fuga de un proveedor del sistema, como la acción correspondiente a dejar de participar activamente en el sistema de Mercado Público durante un periodo de tiempo determinado, según un criterio particular de actividad. Es una definición arbitraria, que depende más que nada del horizonte de inactividad que la dirección determine válido para considerar como fugado, de allí la última expresión de “criterio particular de actividad”.

Existen múltiples criterios para caracterizar y describir fuga. En la dirección se han planteado 2. El primero denominado “criterio de la máxima ventana de inactividad”, criterio propuesto por la división de Mercado Público¹³ y el criterio denominado “Semestre Móvil de Actividad” diseñado específicamente para la realización de esta investigación.

A continuación se describirá el criterio se fuga empleado para caracterizar el problema y modelar en términos reales al fenómeno. Es importante destacar que la fuga se medirá como un estado que depende del tiempo, en este caso se medirá mes a mes. Lo que sigue, es un breve estudio de la fuga actual en el sistema, que permitirá tener una idea más clara del problema que se desea abordar.

Primero, para unificar lenguaje, se entregarán las definiciones de estado de un proveedor que se emplearán en el estudio

2.2.2. Diagrama de Estados de un proveedor en el Sistema

Para caracterizar el comportamiento de un proveedor en el sistema se han diseñado tres estados que describen su dinámica. Se asignará a cada proveedor un “estado” mensual, según su comportamiento transaccional. La caracterización de estado contempla la distinción entre inactivo y fugado transaccional, pues no necesariamente al estar inactivo un mes, se está fugado (el criterio Activo incluye a 2 estados).

- **PROVEEDOR REGISTRADO:** Proveedor logeado en Mercado Público, pero aún no presenta ofertas ni adjudicaciones (stand by).
- **PROVEEDOR TRANSANDO:** Proveedor que presenta al menos una Oferta u Orden de Compra en un instante de tiempo determinado.

¹³ Criterio diseñado en el contexto de otro estudio, pero que se decide mencionar aquí de modo comparativo pues es una alternativa a la modelación de fuga. En el anexo D se describe este criterio en mayor profundidad

- **PROVEEDOR INACTIVO (NO TRANSANDO):** Proveedor que NO presenta ofertas ni Órdenes de Compra en un instante de tiempo (inactivo en transacción).

Existen criterios para catalogar a los proveedores. Actualmente se utiliza el criterio de actividad según tiempo, es decir un proveedor está activo si presenta transacciones en un intervalo de tiempo (por ejemplo un año).

Se propone una distinción de criterios para caracterizar la fuga:

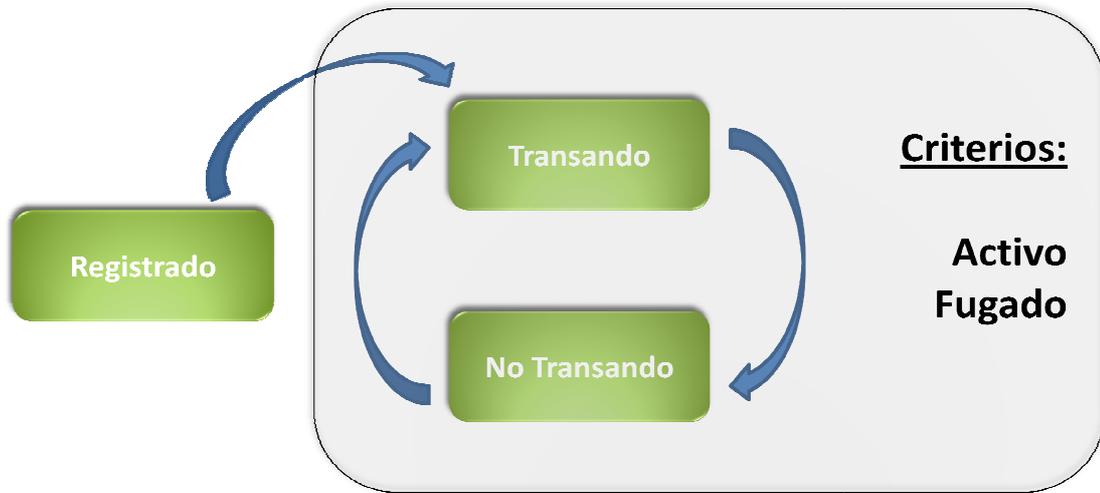
- **PROVEEDOR ACTIVO:** Proveedor que en un intervalo de tiempo **definido** se encuentra en el estado “transando” o en el estado “no transando” y no cumple criterio de fuga transaccional.
- **PROVEEDOR FUGADO (TRANSACCIONAL):** Proveedor inactivo que cumple con condiciones de criterio de fuga transaccional.

Es importante considerar que un proveedor catalogado como “fugado transaccional” puede seguir logeándose en el sistema, pero para efectos de los objetivos estratégicos de la Dirección ChileCompra esto no reviste mayor relevancia, pues el foco es que el proveedor logre efectivamente transar con el estado.

Por simplicidad, se referirá al criterio de estado “fugado transaccional” como “fugado”, pudiendo emplearse otras distinciones como por ejemplo “no operativo”.

A través del siguiente grafo se puede observar la dinámica de un proveedor en el sistema.

Ilustración 7: Grafo de estados de un proveedor en el sistema



Se muestran los estados y los criterios asociados

(Fuente: Elaboración Propia)

2.2.3. Criterio de fuga: “Semestre Móvil de Actividad” (SMA)

El criterio del Semestre Móvil de Actividad intenta capturar la actividad en torno a semestres móviles. En rigor este criterio cataloga como fugado a un proveedor cuando este deja de participar activamente en la plataforma por 6 meses seguidos hacia atrás incluyendo el mes corriente, es decir 6 meses después de su última transacción en el sistema.

Ilustración 8: Criterio de Fuga “Semestre Móvil de Actividad” (SMA)

Prov	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9
A	0	0	1	0	0	0	0	1	0
B	0	0	0	0	1	0	1	0	0
C	1	1	0	0	0	0	0	0	0

(Fuente: Elaboración Propia)

En la ilustración anterior, el proveedor C es analizado en 3 oportunidades. Se observa que presenta actividad en los meses 1 y 2, por ende en los meses 6 y 7 aún es considerado activo. En cambio, ya en el mes 8, el semestre móvil no posee ningún rango de actividad, por ende en este caso el proveedor será considerado como fugado.

Se emplea como barrera el semestre pues se puede demostrar que actualmente el periodo móvil de 6 meses describe adecuadamente la dinámica de fuga de los

proveedores. Si se consideran los periodos *inactivos dentro del rango activo se obtiene que el 76,3%* de los proveedores posee ventanas iguales o menores a 6 meses. Por otro lado, *la probabilidad de encontrar ciclos de 6 meses en proveedores es menor al 2,1%. Por ende, evaluar actividad en intervalo de semestre móvil es un buen indicador y candidato a predictor*¹⁴.

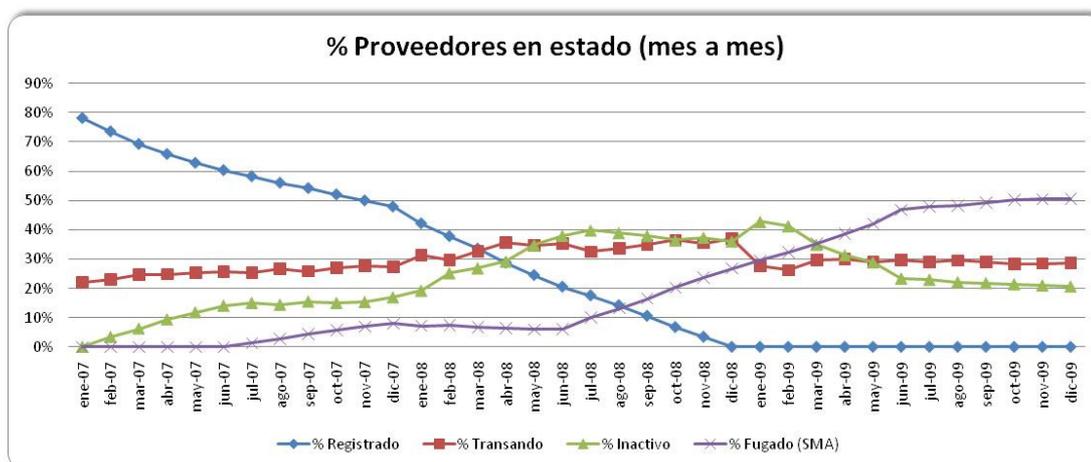
2.2.4. Análisis de la fuga en el sistema

Para caracterizar la fuga se han tomado 2 muestras de datos de proveedores y sus respectivos indicadores de actividad mensual. Se creó una base de datos en donde se describe el nivel de transacciones mes a mes de los proveedores que fueron considerados activos durante 2008 y 2009.

En el caso de los proveedores de 2008, se considerarán 81.449 proveedores únicos que entre enero y diciembre de 2008 presentaron al menos una oferta o se les adjudicó una orden de compra. Criterio similar fue empleado en el caso de la base 2009, en donde se consideraron 89.533 registros (esta última base solo se utilizará en una oportunidad a modo comparativo).

En la figura, se observa la evolución de la fuga de proveedores según su estado. Es importante mencionar que **el 57,9% de los proveedores transando en 2008, lo hizo en 2007 y que el 46,8% de los proveedores se considera fugado a junio de 2009, en donde el crecimiento de la fuga se estabiliza.**

Ilustración 9: Proveedores en Estado (Base activos 2008)



(Fuente: Elaboración Propia)

Mayor detalle de la evolución mes a mes es posible observar en la tabla siguiente.

¹⁴ En el anexo E se detallan estos cálculos con gráficas adecuadas para demostrar el criterio seleccionado.

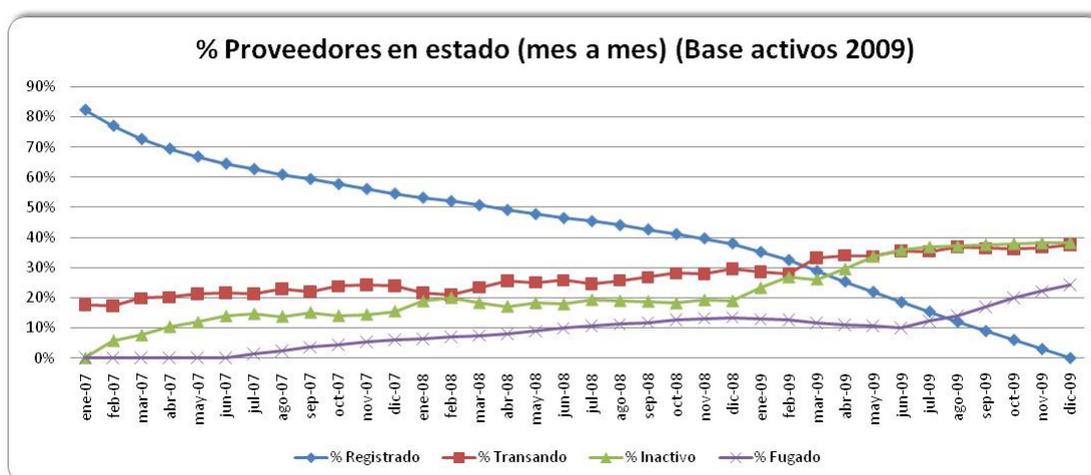
Tabla 2: Evolución de estados en el tiempo (Base activos 2008)

ESTADO	ene-07	jun-07	ene-08	jun-08	ene-09	jun-09	dic-09
Registrado	78,0%	60,2%	42,1%	20,5%	0,0%	0,0%	0,0%
Transando	22,0%	25,6%	31,4%	35,3%	27,6%	29,8%	28,7%
Inactivo	0,0%	14,2%	19,2%	38,1%	42,8%	23,4%	20,7%
Fugado	0,0%	0,0%	7,3%	6,1%	29,5%	46,8%	50,6%

(Fuente: Elaboración Propia)

Al observar el comportamiento según estados en los proveedores base activos 2009, se distingue que el comportamiento es similar, ya que cerca del 60% de los proveedores que transan en 2009, lo hizo antes.

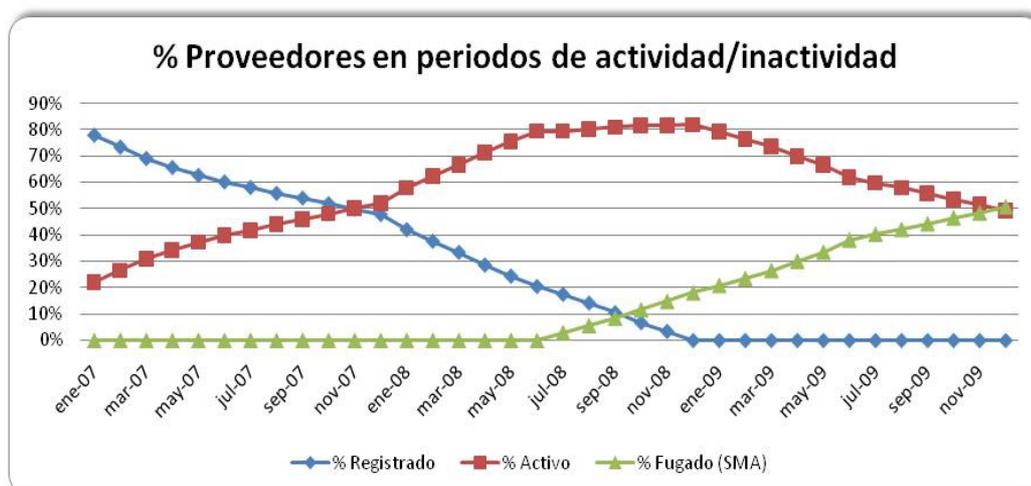
Ilustración 10: Proveedores en Estado. Base activos 2009



(Fuente: Elaboración Propia)

Finalmente la caracterización de estados en el tiempo de los proveedores, evaluando el estado registrado, activos (que transa o que está inactivo por un periodo menor al límite para considerarse fugado) y el estado de inactividad temporal que cumple con criterio SMA, genera el siguiente gráfico de estados y nivel de fuga en el tiempo.

Ilustración 11: Porcentaje de proveedores en estados en el tiempo



(Fuente: Elaboración Propia)

Tabla 3: Evolución de estados en el tiempo, Base activos 2008 (tasa de fuga)

ESTADO	ene-07	jun-07	ene-08	jun-08	ene-09	jun-09	dic-09
Registrado	78,0%	60,2%	42,1%	20,5%	0,0%	0,0%	0,0%
Activo	22,0%	39,8%	57,9%	79,5%	79,2%	61,9%	49,4%
Fugado	0,0%	0,0%	0,0%	0,0%	20,8%	38,1%	50,6%

(Fuente: Elaboración Propia)

A los 6 meses la fuga alcanza un 38,1% de proveedores fugados y a los 12 meses bordea el 50% de los proveedores.

2.2.5. Consecuencias del problema de la fuga en el sistema

Para hacer un correcto análisis se definirán a continuación 2 indicadores empleados en la literatura para evaluar la retención de clientes. Según Huete (Huete, 1996) una empresa fideliza a sus clientes en la medida que logra generar una tasa de retención alta. Define la retención como repetición de compra en un determinado periodo de tiempo. En este caso definiremos la retención como la “repetición de la presentación de una oferta o la adjudicación de una orden de compra en el Mercado Público”. El símil de esta anterior definición es la tasa de deserción, definida como el porcentaje de clientes que no repiten en un periodo determinado. (Macias, 2008).

Ecuación 1: Tasa de Deserción

$$Tasa\ de\ Retención = 1 - \frac{Proveedores\ Fugados}{Proveedores\ Transando\ en\ Mercado\ Público}^{15}$$

(Fuente: "Revitalizando los Servicios" (Huete, 1996))

Ecuación 2: Tasa de Retención

$$Tasa\ de\ Retención = 1 - Tasa\ de\ Deserción$$

(Fuente: "Revitalizando los Servicios" (Huete, 1996))

Por otro lado, Huete en su estudio, afirma que a través de la tasa de deserciones es posible definir la vida media de un cliente con la institución. La vida media de un cliente se obtiene a través de la tasa de deserción expresada en % por unidad de tiempo.

Ecuación 3: Vida media de un cliente con la empresa

$$Vida\ media\ de\ un\ cliente = \frac{1}{Tasa\ de\ deserción}$$

(Fuente: "Revitalizando los Servicios" (Huete, 1996))

En la siguiente tabla se entregan los valores de fuga y tasa de retención asociada

Tabla 4: Indicadores de fidelización de clientes

%Fuga	Fuga Según Criterio	Tasa de Retención
ene-09	20,8%	79,2%
jun-09	38,1%	61,9%
dic-09	50,6%	49,4%

(Fuente: Elaboración Propia)

Se observa que la tasa a 12 meses de retención de proveedores bordea el 50%, esto significa que a los 12 meses el 50% de los proveedores que participaron el año anterior ya no lo están haciendo. Esto podría interpretarse como una "tasa de recambio a 12 meses" del 50%. Se estima que este valor impacta negativamente en el encuentro entre la oferta y la demanda del mercado público, pues existe una cantidad alta de

¹⁵ La definición de tasa de retención (o de fidelización) es extraída de un estudio realizado por el banco ECB (European Central Bank) que trata de un análisis de los beneficios que trae disminuir el porcentaje de clientes fugados anualmente

proveedores que podrían perfectamente participar de los procesos de licitación actuales, ya que poseen el aprendizaje funcional del sistema.

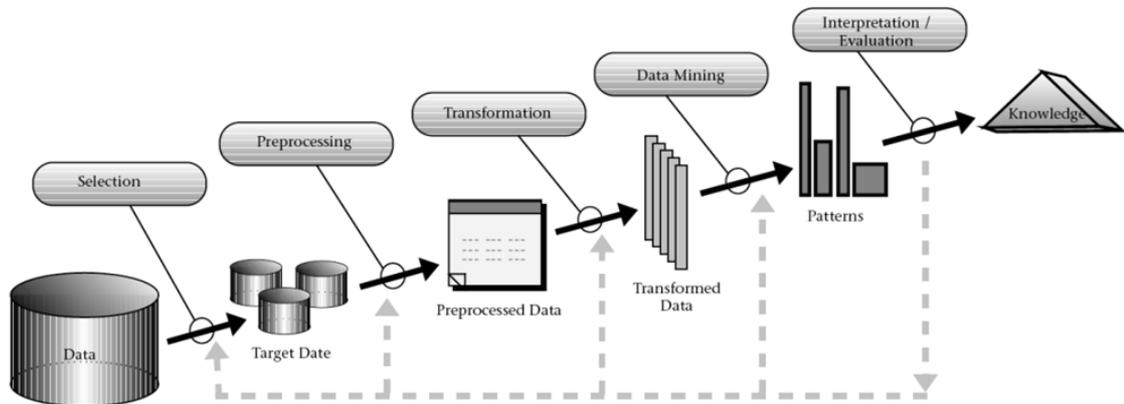
La vida media de un proveedor entonces, podría estimarse que bordea los 2 años. Lo que habla de un fuerte nivel de recambio en la plataforma.

Es importante considerar que a los 6 meses en el caso de la base de 2008 cerca del 15% de los proveedores fugados no presenta una OC adjudicada (equivalente a cerca del 40% de todos los fugados).

3. CAPITULO 3: MARCO CONCEPTUAL

El concepto de minería de datos es una nueva forma de analizar datos para obtener información. En rigor la minería de datos en la literatura es concebida como una parte de un proceso de mayor envergadura denominado proceso *Knowledge Discovery in Databases* (KDD) cuyos pasos principales se muestran en la figura siguiente.

Ilustración 12: Fases de un proceso KDD



(Fuente: (From data mining to knowledge discovery in databases, 1996))

Según Fayyad (From data mining to knowledge discovery in databases, 1996) el proceso KDD es *"The nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data"*, entendiendo el concepto de data mining (minería de datos) como una parte de él.

En la literatura es posible encontrar una multiplicidad de definiciones para el término minería de datos, entre las cuales destacan las definiciones de Hand (Hand, y otros, 2001) y Witten (Witten, y otros, 2005).

Hand indica que *"data mining is the process of secondary analysis of large databases aimed at finding unsuspected relationships which are of interest or value to the database owners"* (Hand, y otros, 2001). Por otro lado, Witten se refiere a la minería de datos como *"a practical topic and involves learning in a practical, not theoretical sense"* consistente en *"techniques for finding and describing structural patterns in data as a tool for helping to explain that data and make predictions from it"* (Witten, y otros, 2005).

Estas dos definiciones escogidas definen a la minería de datos principalmente como un **proceso de análisis** que intenta buscar **relaciones desconocidas** entre

datos. Se habla de un **proceso práctico**, que ayuda a explicar fenómenos y realizar **predicciones** con ellos.

Entendiendo la minería de datos como proceso, se requiere una metodología para enfrentarlo. La cual se describirá a continuación.

3.1. Metodología de Minería de datos

A continuación se propone lo que será la metodología para llevar a cabo el desarrollo del estudio. Todo proyecto de Data Mining requiere la aplicación de una cierta metodología. Actualmente existen varias metodologías predeterminadas que permiten un adecuado control y planificación del proyecto en el tiempo y facilitan el desarrollo del mismo.

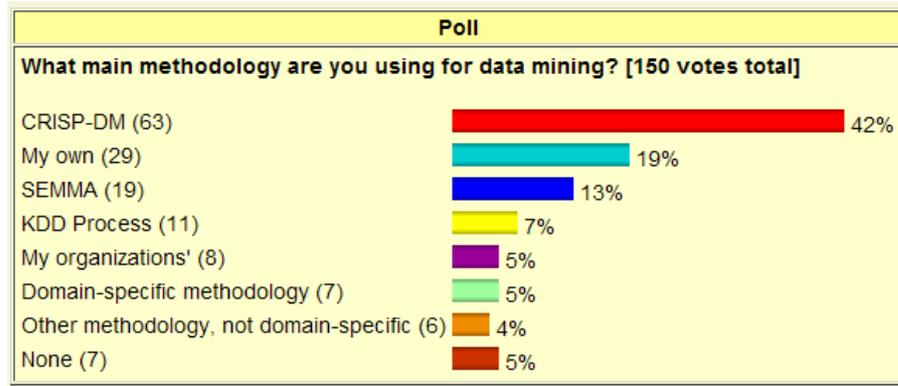
Entre las metodologías disponibles en la literatura podemos encontrar las siguientes.

- Metodología CRISP-DM (Cross Industry Standard Process of Data Mining) (CRISP-DM, 2010).
- Metodología SEMMA (Sample, Explore, Modify, Model, Asses), (SEMMA, 2010)
- Metodología SQL SERVER 2005 (Tang, y otros, 2005)
- Entre otras.

Para la selección de la metodología lo más recomendable es realizar una comparación entre las metodologías planteadas anteriormente. Pero luego de investigar en la literatura, se ha encontrado una encuesta realizada por el reconocido portal para el análisis de datos KDNuggets (KDNuggets, 2010), en donde se consultó a expertos en el tema acerca de que metodologías emplean para enfrentar procesos de minería de datos.

Los resultados arrojaron que CRISP-DM es la metodología más empleada, seguido por la alternativa "propia". Por ende, siguiendo este esquema, se planifica utilizar la metodología CRISP-DM y realizar alguna variación al proceso si fuese estrictamente necesario, con el fin de adecuarla al problema particular.

Ilustración 13: Encuesta sobre metodologías empleadas en proyectos de Data Mining

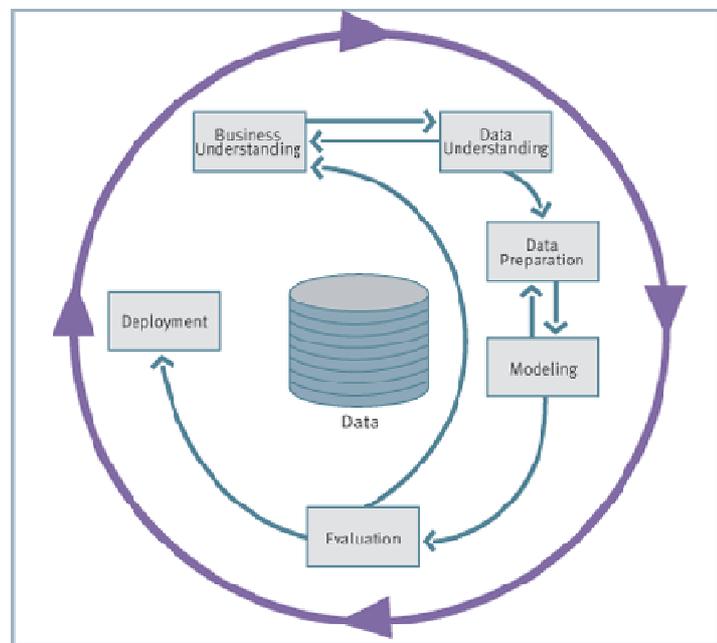


(Fuente: (KDnuggets, 2010))

3.1.1. Metodología CRISP-DM

La metodología CRISP-DM (CRISP-DM, 2010) está descrita en términos de un modelo de proceso. La visión general del ciclo de vida de un proyecto de minería de datos, consiste en seis fases, las cuales se representan en el círculo de la figura. Es posible que existan ciclos, debido a la naturaleza de los proyectos trabajados, por eso se ilustra como un ciclo en constante dinamismo. Actualmente esta metodología se encuentra en su versión 1.0. Es una metodología abierta y disponible para su uso en diversas aplicaciones.

Ilustración 14: Fases del proceso CRISP-DM



(Fuente: (CRISP-DM, 2010))

En general el modelo aborda las siguientes etapas, las cuales se describen de esta misma forma en su plataforma oficial.

Análisis del problema: Fase inicial enfocada a entender los objetivos y requerimientos desde una perspectiva de negocio; para luego definirlos en términos de un problema de Minería de Datos y diseñar un plan para satisfacerlos.

Comprensión de los datos: Se hace una recolección y exploración inicial de los datos para familiarizarse con ellos e identificar problemas de calidad. Además, se trata de descubrir o estimar las relaciones más evidentes para formular las primeras hipótesis sobre información oculta en ellos.

Preparación de los datos: Esta fase cubre todas las actividades necesarias para construir la colección de datos que finalmente será minada a partir del grupo inicial. Incluye la colección, exploración, limpieza, transformación y construcción de datos.

Modelado: Durante esta fase se aplican varias técnicas de modelado. Comúnmente existen varias técnicas para resolver un problema de Minería de Datos del mismo tipo. Incluye la evaluación desde el punto de vista de precisión de los modelos.

Evaluación: Al llegar a esta fase se tendrán los modelos de mayor calidad desde la perspectiva de la precisión. Se impone una evaluación de los modelos y de los pasos que se siguieron para su construcción, a fin de determinar si responden apropiadamente a los objetivos de negocio que se determinaron en la primera fase. Es de vital importancia analizar si alguna regla del negocio no fue tomada en cuenta con el suficiente peso.

Desarrollo: En dependencia de los requerimientos y objetivos, la fase de desarrollo puede ser tan simple como generar un reporte, o tan compleja como emprender un proceso de mayor envergadura. En ocasiones, son los clientes y no los desarrolladores quienes implementan esta fase; deben comprender cómo desarrollarla. Se hace imprescindible documentar y presentar los resultados de manera que todos los puedan entender.

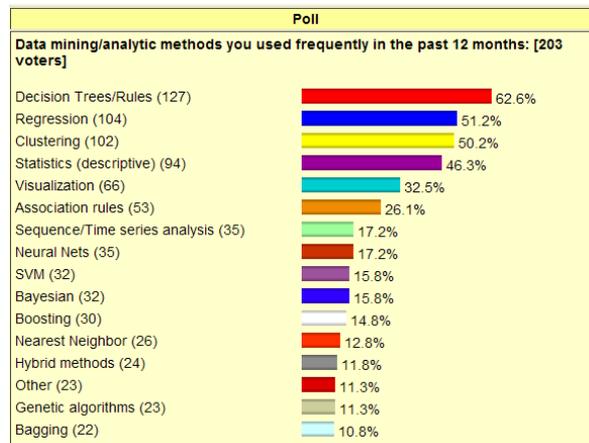
3.2. Técnicas de Minería de datos

En general, cualquiera sea el problema a resolver mediante procesos de Minería de Datos, nunca existe una única técnica para solucionarlo, sino que puede ser abordado por más de una técnica (Aluja, 2001). El número de técnicas es muy grande y sólo puede crecer en el futuro, según expertos en el tema¹⁶, actualmente existen más

¹⁶ Reunión con el profesor del DII. Richard Weber

de 17 técnicas puras. Según KDnuggets las técnicas más empleadas para el análisis de datos son las de tipo árboles de decisión y modelos de regresión, según como se observa en el siguiente gráfico.

Ilustración 15: Técnicas utilizadas para realizar Minería de Datos (2007)



(Fuente: (KDnuggets, 2010))

Las técnicas de minería de datos se clasifican en dos grandes categorías: supervisadas o predictivas y no supervisadas o descriptivas (Weiss, y otros, 1998). En la siguiente figura (Molina, y otros, 2004) se ilustran los tipos de técnicas existentes clasificadas por categoría.

Ilustración 16: Técnicas de Minería de Datos



(Fuente: "Técnicas de Análisis de Datos" (Molina, y otros, 2004))

Según Molina y García (Molina, y otros, 2004), una técnica constituye el enfoque conceptual para extraer la información de los datos, y, en general es implementada por

varios algoritmos. Cada algoritmo representa, en la práctica, la manera de desarrollar una determinada técnica paso a paso. Se requiere entonces, para escoger la técnica más apropiada para cada problema conocer los algoritmos y sus aplicaciones.

En general, las predicciones son empleadas para prever comportamientos, mientras que una descripción ayuda a la comprensión de los mismos. De esta forma, hay algoritmos que sirven para ambos tipos de análisis, pues es posible que una predicción se vuelva una descripción y viceversa.

Para tomar la decisión de qué técnicas escoger se han estudiado diversas aplicaciones de minería de datos en distintos campos. La idea es recoger de manera práctica las capacidades de cada algoritmo.

Los estudios que han sido consultados, son los siguientes:

- “Data Mining en la Empresa y en las Finanzas utilizando Tecnologías Inteligentes” (Weber, 2000).
- “Predicción de Fugas de Clientes para una institución financiera mediante Support Vector Machines” (Miranda, y otros, 2005).
- “Técnicas de Minería de Datos para la Retención de Clientes en el Sector Asegurador” (Macias, 2008)
- “Explotación de Información Aplicada a Inteligencia Criminal en Argentina” (Britos, y otros, 2008)
- “Modelo de Predicción de Fuga de Clientes para Banco Estado” (Briones, 2002)
- “Caracterización de la Población Carcelaria en Argentina Mediante la Aplicación de Minería de Datos para la Prevención de Hechos Delictivos” (Gutierrez, 2008).
- “Empleo de Minería de Datos en la predicción de Diabetes” (Acosta, y otros, 2008)

Las principales fuentes bibliográficas en este capítulo corresponden al libro de José Manuel Molina López y Jesús García Romero, llamado “Técnicas de análisis de datos: Aplicaciones prácticas utilizando Microsoft Excel y Weka”, (Molina, y otros, 2004), el libro, “Data Mining: Practical Machine Learning tolos and techniques”, de Ian H. Witten y Eibe Frank (Witten, y otros, 2005) y el libro “Weka manual for versión 3.61” de la Universidad de Waikato (Waikato, 2010).

A continuación se describirán las principales técnicas de minería de datos que se utilizarán durante este estudio. Además se describirán conceptualmente los algoritmos empleados en cada caso.

3.2.1. Clustering

Técnica, también llamada de agrupamiento, permite la identificación de segmentos o grupos donde los elementos poseen similitud entre sí. Así se puede segmentar el conjunto de datos en función a las características de los mismos.

La principal característica de esta técnica es el uso de una medida de similitud que, en general, está basada en los atributos que describen a los objetos, y se define por proximidad en un espacio multidimensional.

Entre las medidas de similitud destaca la distancia euclideana empleada por ejemplo en el algoritmo de k-medias.

Ecuación 4: Medida de distancia euclideana

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$$

(Fuente: "Técnicas de Análisis de Datos" (Molina, y otros, 2004))

3.2.1.1. Algoritmo K-medias

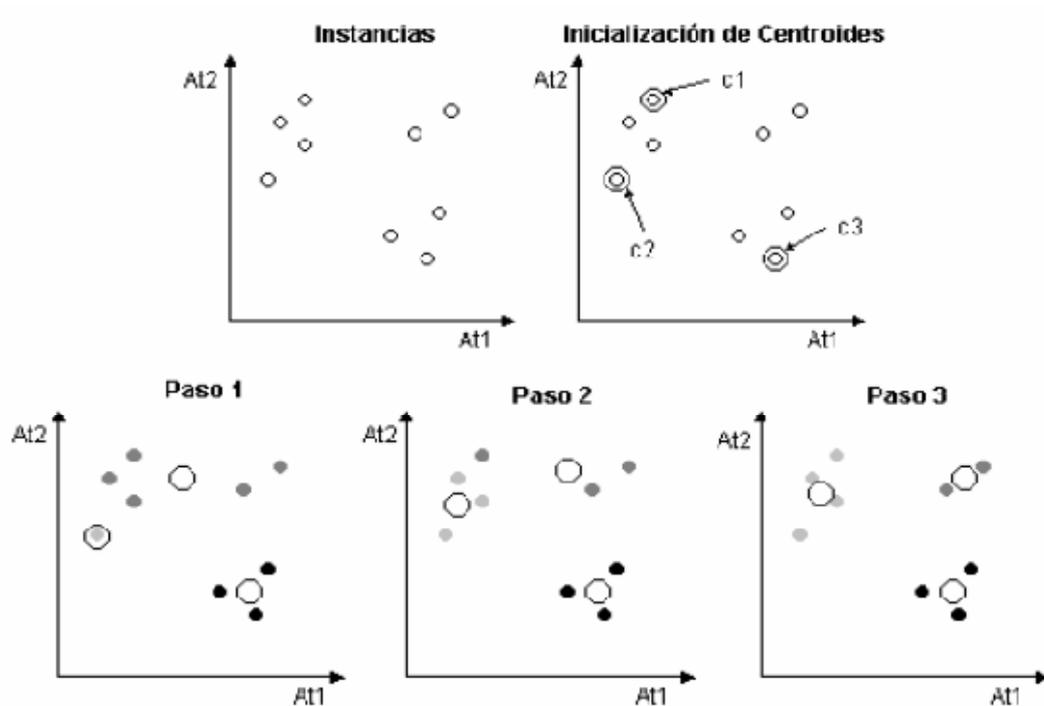
Uno de los algoritmos más utilizados para hacer clustering es el k-medias (Witten, y otros, 2005). Se caracteriza por que requiere que se le indique cuantos clusters o segmentos se quieren encontrar con los datos analizados (de allí su nombre "k"). Por otro lado, para cada una de las instancias, el algoritmo calcula una medida de distancia (por defecto se utiliza la distancia Euclídea, pero puede ser otro tipo de métrica) que le separa del centroide definido inicialmente con k elementos. Este es el paso que le da la segunda parte del nombre, pues se calcula la media "mean" o la moda "mode" según se trate de atributos numéricos o simbólicos.

En general el algoritmo se puede resumir en los siguientes pasos.

- Seleccionar el k, que determina al número de elementos inicial para la generación de clústeres (semillas).
- Para cada elemento siguiente, asignarle la clase más parecida a las seleccionadas inicialmente.
- Calcular nuevamente los Centroides, que se convierten en las nuevas semillas.
- Si no se llega a un criterio de convergencia (por ejemplo, dos iteraciones no cambian las clasificaciones), volver al paso 2.

En la siguiente figura, extraída de (Molina, y otros, 2004) se muestra como ocurre un proceso de clustering.

Ilustración 17: Proceso de clustering con dos atributos.



(Fuente: "Técnicas de Análisis de Datos" (Molina, y otros, 2004))

3.2.1.2. Algoritmo EM

Este es otro algoritmo de clustering, pero su lógica difiere a la de k-medias. Desde este punto de vista, lo que se busca es el grupo de clusters más probables dados los datos. Ahora los ejemplos analizados tienen ciertas probabilidades de pertenecer a un cluster. Se tienen entonces, k distribuciones de probabilidad que representan los k clusters.

El algoritmo EM (Expectation Maximization) empieza adivinando los parámetros de las distribuciones (dicho de otro modo, se empieza adivinando las probabilidades de que un objeto pertenezca a una clase) y, a continuación, los utiliza para calcular las probabilidades de que cada objeto pertenezca a un cluster y usa esas probabilidades para re-estimar los parámetros de las probabilidades, hasta converger.

Este algoritmo recibe su nombre de los dos pasos en los que se basa cada iteración: el cálculo de las probabilidades de los grupos o los valores esperados de los grupos, denominado *expectation*; y el cálculo de los valores de los parámetros de las distribuciones, denominado *maximization*, en el que se maximiza la verosimilitud de las distribuciones dados los datos.

En el algoritmo k-medias se finalizaba cuando ningún ejemplo de entrenamiento cambiaba de cluster en una iteración (convergencia). En el algoritmo EM es un poco más complicado, dado que el algoritmo tiende a converger pero nunca se llega a ningún punto fijo. Sin embargo, se puede ver cuánto se acerca calculando la verosimilitud (likelihood) general de los datos con esos parámetros, multiplicando las probabilidades de los ejemplos, según la ecuación:

Ecuación 5: Cálculo de verosimilitud en algoritmo EM

$$\prod_{i=1}^N \left(\sum_j^{cluster} p_j P(x_i|j) \right)$$

(Fuente: "Técnicas de Análisis de Datos" (Molina, y otros, 2004))

En esta ecuación j representa cada uno de los clusters del sistema, y p_j la probabilidad de dicho cluster. La verosimilitud es una medida de lo "bueno" que es el clustering, y se incrementa con cada iteración del algoritmo EM. Se seguirá iterando hasta que dicha medida se incremente un valor despreciable.

El principal uso de este algoritmo en este estudio es que a diferencia de k-medias, EM no requiere obligatoriamente un determinado número de clusters iniciales para generar, sino que el mismo algoritmo entrega el cluster óptimo.

Se evaluará primero el conjunto de datos con el algoritmo EM, obteniendo un número X de clusters, el cual luego se ingresará a los parámetros de k-medias.

3.2.2. Clasificación

La clasificación consiste en dividir un conjunto de datos en grupos mutuamente excluyentes, de tal forma que cada miembro de un grupo esté lo más cerca posible de otros miembros y grupos diferentes estén lo más lejos posible entre ellos donde la distancia se mide con respecto a las variables especificadas, que se quieren predecir.

Las técnicas de clasificación también son empleadas para predecir, pues pueden predecir en donde le corresponde estar a una instancia que no haya sido clasificada previamente.

Existen varias técnicas de clasificación. Se han estudiado varias, incluyendo redes bayesianas, arboles de decisión, etc. Finalmente se ha optado por los arboles de decisión y una técnica específica, que se describirá a continuación.

3.2.2.1. Árboles de Decisión

Si bien, el primer sistema que construía árboles de decisión fue CLS de Hunt, desarrollado en 1959, en 1979 Quinlan desarrolla el sistema ID3 (Quinlan, 1979), que él denominaría simplemente herramienta porque la consideraba experimental. La versión definitiva, presentada por su autor Quinlan como un sistema de aprendizaje, es el sistema C4.5 que expone con cierto detalle en la obra *C4.5: Programs for Machine Learning* (Quinlan, 1993). La evolución comercial de ese sistema es otro denominado C5 del mismo autor.

Un árbol de decisión puede interpretarse esencialmente como una serie de reglas compactadas para su representación en forma de árbol.

El ID3 es capaz de tratar con atributos cuyos valores sean discretos o continuos. En el primer caso, el árbol de decisión generado tendrá tantas ramas como valores posibles tome el atributo. Si los valores del atributo son continuos, el ID3 no clasifica correctamente los ejemplos dados. Por ello, Quinlan propuso el C4.5.

El algoritmo C4.5 genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente. Antes de realizar una partición de datos, el algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos, seleccionando la prueba que aporte mayor ganancia de información.

Cualquier árbol de decisión se puede convertir en una estructura de reglas de clasificación, asumiendo que esta estructura es del tipo: Si <Condición> Entonces <Clase>. El algoritmo de generación de reglas consiste básicamente en, por cada rama del árbol de decisión estarán sus valores y la etiqueta del nodo hoja tendrá la clasificación. Este procedimiento generaría un sistema de reglas con mayor complejidad de la necesaria (en volumen). Por ello, el sistema C4.5 realiza un podado de las reglas obtenidas, generando un árbol de reglas significativo.

3.2.2.2. Reglas de Clasificación

Si bien existe la posibilidad de convertir el resultado de arboles de decisión en reglas de clasificación, existen algoritmos que lo entregan directamente.

La inducción de reglas se puede lograr fundamentalmente mediante dos caminos: Generando un árbol de decisión y extrayendo de él las reglas, como puede hacer el sistema C4.5 o bien mediante una estrategia de covering, consistente en tener en cuenta cada vez una clase y buscar las reglas necesarias para cubrir (cover) todos los ejemplos de esa clase; cuando se obtiene una regla se eliminan todos los ejemplos que cubre y se continúa buscando más reglas hasta que no haya más ejemplos de la clase.

El Algoritmo PART recibe el nombre por su modo de actuación: “obtaining rules from PARTial decision trees”, y fue desarrollado por el grupo neozelandés que construyó el entorno WEKA.

El sistema se basa en las dos estrategias básicas para la inducción de reglas: el covering y la generación de reglas a partir de árboles de decisión. Adopta la estrategia del covering (con lo que se obtiene una lista de decisión) dado que genera una regla, elimina los ejemplares que dicha regla cubre y continúa generando reglas hasta que no queden ejemplos por clasificar.

3.3. Software de Minería de Datos

Para realizar el Preprocesado de los datos de donde se desea extraer conocimiento, se requiere contar con un arsenal de herramientas Software. A continuación se describirán las principales herramientas de software que se espera utilizar.

Los datos se encuentran en los repositorios de la institución, en donde se trabaja con software administradores de bases de datos. En este caso es Microsoft SQL Server 2008.

Para la exploración de datos, es decir, extraerlos desde las fuentes originales y llevarlos hasta el repositorio principal, es posible apoyarse con Microsoft Excel 2007.

Ahora en lo que se refiere al Software para el análisis de datos, se han descargado diversas herramientas, las cuales contienen una gran cantidad de Algoritmos (técnicas) de Minería de datos. Además incluyen aplicaciones para importar datos de diversas fuentes digitales o gestores de bases de datos reconocidos (Excel, SQL Server, etc.). Estas herramientas, incorporan además, un conjunto de filtros que permiten realizar las operaciones necesarias para la transformación y limpieza de datos.

Se han evaluado diversos software, entre ellos WEKA, RapidMiner y Data Engine. A continuación se describirá uno de ellos.

3.3.1. WEKA

Weka fue desarrollado en la Universidad de Waikato (Waikato, 2010) en Nueva Zelanda. Es un sistema desarrollado en Java, un lenguaje de programación orientado a objetos multiplataforma. Weka funciona para diversos sistemas operativos, Windows, mac, Linux, etc.

Para el desarrollo de este estudio se optó como herramienta el software WEKA, pues es un software de libre distribución (bajo licencia GPL lo cual ha impulsado que

sea una de las suites más utilizadas en el área en los últimos años), caracterizándose por su facilidad de uso y programación (Ambiente con consola Java).

El paquete Weka contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades. Al ejecutar la aplicación nos aparece el selector de interfaz de Weka (Weka GUI Chooser) que da la opción de seleccionar entre cuatro posibles interfaces de usuario para acceder a las funcionalidades del programa, éstas son "Simple CLI", "Explorer", "Experimenter" y "Knowledge Flow".

La interfaz Explorer (Explorador) dispone de varios paneles que dan acceso a los componentes principales del banco de trabajo.

El panel "Preprocess" dispone de opciones para importar datos de una base de datos, de un fichero CSV, etc., y para preprocesar estos datos utilizando los denominados algoritmos de filtrado. Estos filtros se pueden utilizar para transformar los datos (por ejemplo convirtiendo datos numéricos en valores discretos) y para eliminar registros o atributos según ciertos criterios previamente especificados. El panel "Classify" permite al usuario aplicar algoritmos de clasificación a los conjuntos de datos resultantes. El panel "Associate" proporciona acceso a las reglas de asociación aprendidas que intentan identificar todas las interrelaciones importantes entre los atributos de los datos. El panel "Cluster" da acceso a las técnicas de clustering o agrupamiento de Weka como por ejemplo el algoritmo k-medias. El panel "Selected attributes" proporciona algoritmos para identificar los atributos más predictivos en un conjunto de datos. El panel "Visualize" muestra una matriz de puntos dispersos donde cada punto individual puede seleccionarse y agrandarse para ser analizados en detalle usando varios operadores de selección.

El principal panel de weka que se utilizará será el panel de explorer, ya que allí contiene las implementaciones de todos los algoritmos antes mencionados.

Las diferentes configuraciones de cada uno de estos algoritmos, las variables a setear, etc. Por cada algoritmo, se detallan en el Anexo F.

3.3.1.1. Test y Evaluación en Weka

Para el caso de la clasificación weka permite la realización de 4 test distintos.

- **“Use training set”**: Es un test que mide la calidad del clasificador empleado utilizando las mismas instancias en las que ha sido entrenado (generado el modelo).
- **“Supplied set test”**: Se evalúa la calidad del clasificador con una muestra de datos diferentes.

- **“Cross-validation”**: Se evalúa la calidad del clasificador mediante validación cruzada utilizando el número de grupos que se especifiquen.
- **“Percentage Split”**: Se evalúa la calidad del clasificador según lo bien que clasifique a un porcentaje de los datos de entrada. Ese porcentaje es configurable. Por defecto se entrega un 66%.

Para evaluar el modelo se entrega información entre las cuales se destacan las estadísticas, como por ejemplo, el porcentaje de instancias clasificadas correctamente (precisión), el estadístico de kappa, que mide lo bien que se ajusta la predicción a la clase real (un valor igual a 1 representa un ajuste de un 100%), entre otros valores.

Finalmente, se presenta un detalle de exactitud para cada una de las clases y la matriz de confusión. Las medidas de exactitud empleadas son las siguientes:

- **Tasa TP (true positive)**: es la tasa de elementos que fueron clasificados en la clase A con respecto a los que de verdad poseen esa clase. Es decir, que cantidad de la clase ha sido tomada por el modelo. En términos de la matriz de confusión, corresponde al valor del elemento de la diagonal dividido por la suma de la fila que representa la clase respectiva. También es llamada “recall”.
- **Tasa FP (false positive)**: es la tasa de elementos que fueron clasificados en la clase A, pero que no la poseen, con respecto a los que no poseen esa clase.
- **Precisión**: proporción de elementos que efectivamente poseen la clase A, entre todos los que fueron clasificados como clase A.
- **F-Measure**: Es una combinación de la precisión y recall. Su fórmula es:

Ecuación 6: Formula de obtención de F-measure

$$F - Measure = \frac{2 * precision * recall}{precision + recall}$$

(Fuente: “Data Mining: Practical Machine Learning tools and techniques” (Witten, y otros, 2005))

- **Matriz de confusión**: Si se tienen k clases, se define la matriz de confusión como una matriz de k x k, en que el número de instancias clasificadas correctamente es la suma de los elementos de la diagonal, la suma de los demás indica la cantidad de instancias clasificadas incorrectamente. Todas las medidas anteriormente presentadas, pueden ser extraídas directamente de la matriz de confusión.
- **Coeficiente Kappa**: Coeficiente estadístico que se emplea para cuantificar el grado de acuerdo entre los observadores, corrige el factor azar. Cuanto más se acerca al cero más azar hay, es casi decir que el diagnostico se hace echando una moneda al aire.

4. CAPITULO 4: COMPRESION DE LOS DATOS

Para determinar las variables a incluir en el modelo, a continuación se describirán los datos transaccionales con los que se cuenta. Estos datos han sido obtenidos desde diversas fuentes internas, por ejemplo el data mart de usuarios, y la base de datos de consulta de la institución.

Se han tenido que diseñar múltiples consultas en lenguaje SQL¹⁷ para poder extraer los datos y construir el repositorio de información a evaluar.

Posteriormente a la extracción se realizará un proceso de transformación de los datos, creación de nuevas variables mezclando información, etc.

Y por último, se seleccionarán las variables a través de un análisis cualitativo, algoritmos de selección de datos a través del software weka y reuniones con expertos en el tema para garantizar que las variables seleccionadas sean adecuadas para la generación del conocimiento esperado.

4.1. Descripción de los Datos:

Para realizar el análisis se han tomado datos desde 2007 a 2009 para diversas variables de sistema. Dado los datos extraídos se ha tomado la determinación de trabajar directamente con los datos asociados a los proveedores activos durante 2009, por razones de calidad de datos, completitud y consistencia de la información. En la tabla siguiente se describe la cantidad de proveedores de la muestra.

Tabla 5: Línea de base para análisis de datos

Proveedores Base	Base 2009
Proveedores Ofertando	69.673
Proveedores Adjudicando	67.910
Proveedores Adjudicando u Ofertando	89.524

(Fuente: Elaboración propia)

Los datos a extraer se pueden clasificar en 3 tipos

- **Identificación:** Datos que describen al proveedor en sí mismo.
- **Transacción:** Datos de comportamiento transaccional del proveedor en el Sistema de Mercado Público.
- **Sistema:** Datos relacionados con el uso de la plataforma.

¹⁷ SQL: Structured Query Language

Tabla 6: Datos para extraer desde sistemas

Campo	Tipo	Descripción	Tipo de Dato	Tiempo y Frecuencia
Rut	Identificación	Identificador base	Texto	No Aplica
Tamaño Empresa	Identificación	Tamaño de la Empresa según catalogación del SII	Texto	Actualizado a 2009
Antigüedad Ofertas	Identificación	Fecha de la primera oferta presentada por el proveedor	Fecha	No Aplica
Antigüedad Orden de Compra	Identificación	Fecha de la primera Orden de Compra presentada por el proveedor	Fecha	No Aplica
Región Proveedor	Identificación	Regiones en las que tiene sucursal el proveedor	Texto	Actualizado a 2009
Rubros Inscritos	Identificación	Rubros en que el proveedor está inscrito en la plataforma	Texto	Actualizado a 2009
Cantidad de Sucursales	Identificación	Cantidad de sucursales asociadas a un Rut proveedor.	Numérico	Actualizado a 2009
Ofertas por Proveedor	Transacción	Cantidad de Ofertas presentadas por proveedor (en estado enviada y rechazada ¹⁸)	Numérico	Mensual enero de 2007 a diciembre de 2009
Órdenes de Compra	Transacción	Cantidad de Órdenes de Compra recibidas por proveedor	Numérico	Mensual enero de 2007 a diciembre de 2009
Montos Transados	Transacción	Montos Transados en cada Orden de Compra (Expresado en Dólar sin IVA)	Numérico	Mensual enero de 2007 a diciembre de 2009
Adjudicaciones por proveedor	Transacción	Cantidad de procesos de licitación en donde el proveedor ha sido adjudicado	Numérico	Mensual enero de 2007 a diciembre de 2009
Rubros transados por proveedor	Transacción	Cantidad de Rubros en que el proveedor ha efectuado transacciones en el sistema	Numérico	Anual desde 2007 a 2009
ChileProveedores	Sistema	Interacción del proveedor con la plataforma de ChileProveedores	Texto	No Aplica

¹⁸ Un organismo público puede rechazar una oferta si se encontrase fuera de las bases de licitación, o si no adjunta los documentos exigidos según lo estipulan las bases.

Login	Sistema	Cantidad de logeos por proveedor	Numérico	Mensual 2009
Número de Direcciones IP	Sistema	Cantidad de direcciones IP distintas desde donde el proveedor se loguea	Numérico	Mensual 2009
Reclamos ¹⁹	Sistema	Cantidad de reclamos presentados por el proveedor	Numérico	Anual desde 2007 a 2009

(Fuente: Elaboración Propia)

4.2. Set de datos seleccionado

Luego de tres meses de construcción de consultas en lenguaje SQL, se ha logrado extraer la información definida inicialmente. Esta será procesada para obtener las variables que serán utilizadas para la construcción de los modelos.

Al observar la siguiente tabla, se podrá entender la razón de que se escoja exclusivamente la base de datos de proveedores activos en 2009 para efectuar el análisis. Las variables vigencia en ChileProveedores y Cantidad de logeos, solo arrojan registros para el periodo descrito.

Finalmente se obtienen 14 variables de los datos mencionados anteriormente. Estas variables no constituyen en su totalidad información transaccional, pues se incluyen además datos de sistema y de identificación del proveedor. De todos modos en general son variables que se generan en la interacción proveedor-Mercado Público, por ende entran en la lógica de la información generada por la transacción.

Tabla 7: Set de Variables

Variable	Descripción	Tipo de Dato	Tiempo y Frecuencia
Tamaño Empresa	Tamaño de la Empresa según catalogación del SII	Texto	2009
Antigüedad Proveedor	Antigüedad en transacciones de proveedor. Definido como la mínima fecha entre la primera oferta y primera OC	Numérico	Con respecto a 2008 y 2009
Estado Proveedor	Estado de un proveedor en el tiempo	Texto	Mensual: Enero 2007 a diciembre 2009
Cantidad de Ofertas	Cantidad de Ofertas Presentadas por proveedor	Numérico	Mensual: Enero 2007 a diciembre 2009

¹⁹ En este caso un reclamo enviado por un proveedor será considerado como una variable de la interacción del proveedor con el sistema.

Cantidad de Ofertas rechazadas	Cantidad de ofertas presentadas que han sido rechazadas por un comprador	Numérico	Mensual: Enero 2007 a diciembre 2009
Cantidad de Órdenes de Compra	Cantidad de Órdenes de Compra recibidas por proveedor	Numérico	Mensual enero de 2007 a diciembre de 2009
Cantidad de Monto Transado	Montos Transados en cada Orden de Compra (Expresado en Dólar sin IVA)	Numérico	Mensual enero de 2007 a diciembre de 2009
Cantidad de Adjudicaciones	Cantidad de Adjudicaciones por proveedor	Numérico	Mensual enero de 2007 a diciembre de 2009
Cantidad de Reclamos	Cantidad de reclamos presentados a la plataforma	Numérico	Anual desde 2007 a 2009
Registro en ChileProveedores	Variable que identifica si el proveedor está o no registrado en ChileProveedores	Booleano	Desde 2006 en adelante
Vigencia en ChileProveedores	Variable que identifica si el proveedor está vigente en ChileProveedores	Booleano	2009
Cantidad de logeos	Cantidad de veces que el proveedor se logea en el sistema	Numérico	2009
WinRate	Cuociente entre las adjudicaciones y las ofertas enviadas. Es una medida del éxito del proveedor en licitaciones.	Numérico	Anual desde 2007 a 2009
OfRate	Cuociente entre las ofertas rechazadas y las ofertas enviadas. Es una medida de la calidad de las ofertas de los proveedores	Numérico	Anual desde 2007 a 2009

(Fuente: Elaboración Propia)

Para efectos del análisis, se excluyen las distinciones de transacciones por tipo (por ejemplo licitaciones públicas y/o tratos directos), debido a que representan un mayor nivel de detalle y para efectos de este estudio se plantea analizar la dinámica general.

Se entenderá que si existen proveedores participando activamente de los convenios marco estos tendrán una relación mayor de órdenes de compra por sobre las ofertas presentadas (por ello la tasas de éxito se define como adjudicaciones sobre ofertas y no órdenes de compra sobre ofertas).

4.3. Limpieza y transformación de la Información

Durante la recopilación de los datos se pudieron observar problemas en la calidad de los mismos, y a su vez, se encontraron varios campos que presentaban

datos nulos o ausentes. En general han sido considerados datos íntegros y se han descartado los que posean cierto grado de desconfianza ya sea por estar “sucios” o por no contar con toda la información necesaria.

Existen algunas variables que no poseen sus valores en 2008 (y 2007), esto se debe a que la plataforma ha ido en constante evolución en lo que se refiere a su modelo de datos, por ende existen valores nulos para cierta cantidad de proveedores (por ejemplo la fecha de primer login ha sido omitida del análisis pues no se cuenta con información fidedigna que sustente su incorporación al data set).

Por otro lado, en el siguiente apartado se explorarán los datos. Allí se describirá si existen anomalías en la información y como serán tratadas desde el punto de vista de no alterar la consistencia de la información.

Primero que todo, se definirá una correcta línea de base para la exploración de los datos. Para ello se propone eliminar proveedores con término de giro, proveedores de nacionalidad extranjera (con ruts en otros formatos o simplemente sin rut). En la siguiente tabla se encuentran los datos que se cruzarán con la línea de base original.

Tabla 8: Proveedores que serán eliminados del estudio

Base	Cantidad
Proveedores Extranjeros	1.978
Proveedores Término Giro	5.982

(Fuente: Elaboración Propia)

El procesamiento a través de cruce de datos arroja el siguiente resultado:

Tabla 9: Preprocesado línea de base

Bases	Base 2009
Base Inicial (Sucia)	89.524
Cruce Extranjeros	187
Cruce Termino Giro	222
Otros	61
Base final	89.053

(Fuente: Elaboración Propia)

En la base han sido eliminados 187 ruts de proveedores extranjeros y 222 con término de giro, junto a 61 ruts en formatos desconocidos (posiblemente erróneos). La base a explorar consta 89.053 para los proveedores activos en 2009.

Ahora analizando la calidad de la información obtenida en la base de datos final se obtiene que existen 6.606 instancias de proveedores con al menos un dato erróneo o fuera de rango. En la siguiente tabla, se explicitan los porcentajes totales de los errores de datos registrados y sus causas para el set de datos final.

Se opta por eliminar estos registros, en vez de rellenar con valores. La decisión pasa por qué comparando las distribuciones con y sin estos valores la diferencia promedio es menor al 1%.

Tabla 10: Dataset final y sus porcentajes de calidad de datos

Variable	Datos Completos	Datos Erróneos	Observación
Estado Transaccional	100,0%	0,0%	
Tamaño Empresa	98,7%	1,3%	Tamaño Empresa: SINDATO
Año Primera Transacción	99,9%	0,1%	Fechas fuera de rango (1900)
Cantidad Adjudicaciones 2009	100,0%	0,0%	
Cantidad Órdenes de Compra 2009	100,0%	0,0%	
Cantidad Ofertas 2009	100,0%	0,0%	
Monto Transado 2009	99,09%	0,01%	Montos con valor negativo
Cantidad Ofertas rechazadas 2009	100,0%	0,0%	
Promedio logeo mensual	100,0%	0,0%	Valores fuera de Rango (>10.000)
Cantidad de meses logeado	100,0%	0,0%	
Registro en ChileProveedores	100,0%	0,0%	
Vigencia en ChileProveedores	100,0%	0,0%	
Reclamos 2009	100,0%	0,0%	
Tasa Of-Rate	100,0%	0,0%	
Tasa Win rate	93,9%	6,1%	Valores fuera de Rango (>100%). La razón es que existen procesos que vienen del año anterior y que son cerrados en 2009.

(Fuente: Elaboración Propia)

Finalmente al eliminar los registros que presentan anomalías se cuenta con una base de 82.446 registros limpios de proveedores, los cuales serán sujetos de análisis y descritos a continuación.

Existen variables que han sido transformadas desde datos numéricos a frecuencias. Esto se debe a que así se minimiza la variabilidad de los datos y se obtiene mayor claridad de los resultados.

4.4. Exploración de las variables

Estado Proveedores: Variable objetivo. Para efectos del análisis los proveedores transando e inactivos están dentro de su periodo activo (es decir el 74,8% de los proveedores a diciembre de 2009 no está considerado fugado).

Tabla 11: Variable: Estado Proveedores

Estado Proveedores	Cantidad	Porcentaje
Transando	29.979	36,4%
Inactivo	31.652	38,4%
Fugado	20.815	25,2%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 18: Variable: Estado Proveedores



(Fuente: Elaboración Propia)

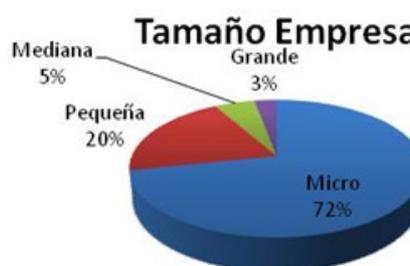
Tamaño Empresa: Se emplea la convención del SII definiendo tamaño de empresa por volumen de ventas. Este dato es extraído del formulario de registro de proveedores.

Tabla 12: Variable: Tamaño Empresa

Tamaño Empresa	Cantidad	Porcentaje
Micro	59.140	71,7%
Pequeña	16.548	20,1%
Mediana	4.344	5,3%
Grande	2.414	2,9%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 19: Variable: Tamaño Empresa



(Fuente: Elaboración Propia)

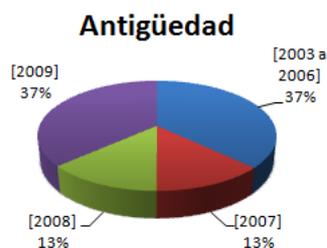
Antigüedad en Transacción: Corresponde al año en que se efectúa la primera transacción en el sistema. Se observa que el 63,4 % corresponde a proveedores que vienen de años anteriores. Se unen los años 2003 a 2006 pues no se tiene la efectiva certeza del año, ya que no se cuenta con los registros asociados.

Tabla 13: Variable: Antigüedad en Transacción

Antigüedad	Cantidad	Porcentaje
[2003 a 2006]	30.626	37,1%
[2007]	10.633	12,9%
[2008]	10.972	13,3%
[2009]	30.215	36,6%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 20: Variable: Antigüedad en Transacción



(Fuente: Elaboración Propia)

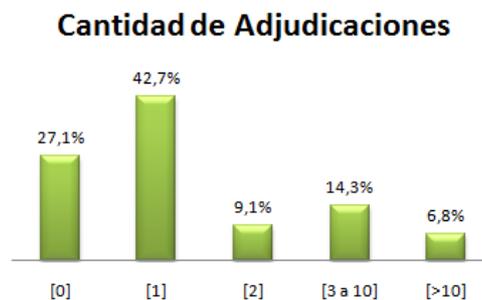
Cantidad de Adjudicaciones: Corresponde a la cantidad de licitaciones en donde al menos una línea de licitación ofertada fue exitosa en 2009.

Tabla 14: Variable: Cantidad de Adjudicaciones

Cantidad de Adj.	Cantidad	Porcentaje
[0]	22.381	27,1%
[1]	35.168	42,7%
[2]	7.486	9,1%
[3 a 10]	11.810	14,3%
[>10]	5.601	6,8%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 21: Variable: Cantidad de Adjudicaciones



(Fuente: Elaboración Propia)

Cantidad de Órdenes de Compra: Corresponde a la cantidad de órdenes de compra recibidas y aceptadas por el proveedor en 2009.

Tabla 15: Variable: Cantidad de Órdenes de Compra

Cantidad de Órdenes de Compra	Cantidad	Porcentaje
[0]	21.062	25,5%
[1]	22.493	27,3%
[2 a 4]	16.141	19,6%
[5 a 15]	13.104	15,9%
[>15]	9.646	11,7%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 22: Variable: Cantidad de Órdenes de Compra



(Fuente: Elaboración Propia)

Cantidad de Ofertas: Corresponde a la cantidad de ofertas enviadas por el proveedor, que figuran en estado aceptada y rechazada (se excluyen las ofertas guardadas)

Tabla 16: Variable: Cantidad de Ofertas

Cantidad de Ofertas	Cantidad	Porcentaje
[0]	19.483	23,6%
[1]	21.116	25,6%
[2 a 4]	18.522	22,5%
[5 a 15]	12.649	15,3%
[>15]	10.676	12,9%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 23: Variable: Cantidad de Ofertas



(Fuente: Elaboración Propia)

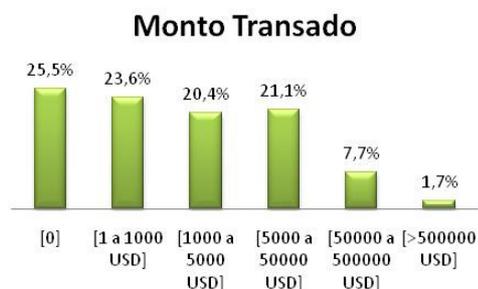
Monto Transado: Corresponde a la suma de los montos asociados a las Órdenes de Compra Recibidas por proveedor.

Tabla 17: Variable: Monto Transado

Monto Transado	Cantidad	Porcentaje
[0]	21.063	25,5%
[1 a 1.000 USD]	19.460	23,6%
[1.000 a 5.000 USD]	16.823	20,4%
[5.000 a 50.000 USD]	17.357	21,1%
[50.000 a 500.000 USD]	6.374	7,7%
[>500.000 USD]	1.369	1,7%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 24: Variable: Monto Transado



(Fuente: Elaboración Propia)

Cantidad de Ofertas Rechazadas: Corresponde a la suma de ofertas enviadas por un proveedor, que han sido rechazadas por el comprador por estar fuera de base o no cumplir con los requisitos mínimos estipulados en las bases de licitación.

Tabla 18: Variable: Cantidad de Ofertas Rechazadas

Oferta Rechazada	Cantidad	Porcentaje
[0]	63.923	77,5%
[1]	9.709	11,8%
[2 a 4]	5.641	6,8%
[5 a 15]	2.258	2,7%
[>15]	915	1,1%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 25: Variable: Cantidad de Ofertas Rechazadas



(Fuente: Elaboración Propia)

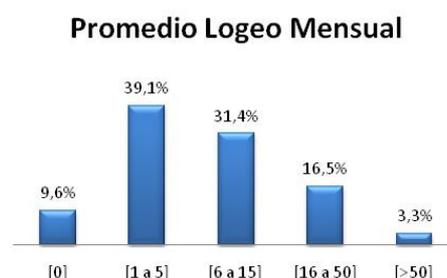
Promedio Logeo Mensual: Corresponde al promedio de la cantidad de veces que el proveedor se logea en Mercado Público, en los meses en donde efectivamente lo hace.

Tabla 19: Variable: Promedio Logeo

Promedio Logeo Mensual	Cantidad	Porcentaje
[0]	7.944	9,6%
[1 a 5]	32.266	39,1%
[6 a 15]	25.871	31,4%
[16 a 50]	13.626	16,5%
[>50]	2.739	3,3%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 26: Variable: Promedio Logeo



(Fuente: Elaboración Propia)

Registro en ChileProveedores: Corresponde al indicador de si la empresa está registrada o no en el Registro de ChileProveedores.

Tabla 20: Variable: Registro en ChileProveedores

Registro en ChileProveedores	Cantidad	Porcentaje
NO	36.987	44,9%
SI	45.459	55,1%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 27: Variable: Registro en ChileProveedores



(Fuente: Elaboración Propia)

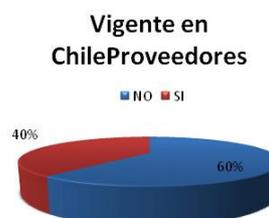
Vigencia en ChileProveedores: Corresponde al indicador de si la empresa está vigente en el Registro de ChileProveedores.

Tabla 21: Variable: Vigente en ChileProveedores

Vigente en ChileProveedores	Cantidad	Porcentaje
NO	49.824	60,4%
SI	32.622	39,6%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 28: Variable: Vigente en ChileProveedores



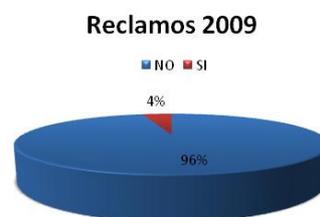
(Fuente: Elaboración Propia)

Reclamos: Indica si en el último año la empresa ha presentado algún tipo de reclamo asociado a su interacción con el portal.

Tabla 22: Variable: Reclamos

Reclamos 2009	Cantidad	Porcentaje
NO	79.084	95,9%
SI	3.362	4,1%
TOTAL	82.446	100,0%

Ilustración 29: Variable: Reclamos



(Fuente: Elaboración Propia)

(Fuente: Elaboración Propia)

Cantidad de Meses Logeado: Corresponde a la cantidad de meses en el año en que el proveedor se logeo en el sistema.

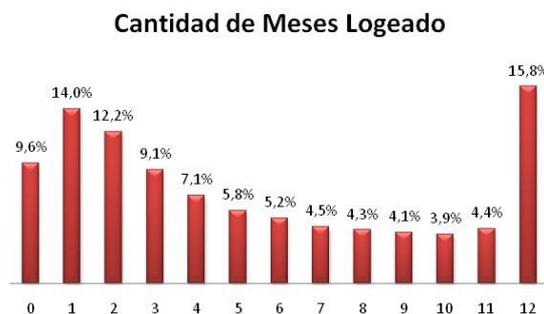
Tabla 23: Variable: Cantidad de Meses Logeado

Cantidad de Meses Logeado	Cantidad	Porcentaje
0	7.944	9,6%
1	11.514	14,0%
2	10.055	12,2%
3	7.505	9,1%
4	5.815	7,1%
5	4.817	5,8%
6	4.312	5,2%

Cantidad de Meses Logeado	Cantidad	Porcentaje
7	3.748	4,5%
8	3.518	4,3%
9	3.349	4,1%
10	3.226	3,9%
11	3.626	4,4%
12	13.017	15,8%
TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Ilustración 30: Variable: Cantidad de Meses Logeado



(Fuente: Elaboración Propia)

Tasa Of-Rate: Corresponde tasa de ofertas rechazadas sobre las ofertas enviadas. Es una medida de la calidad de las ofertas presentadas por el proveedor.

Tabla 24: Variable: Tasa Of-Rate

Tasa Of-Rate	Cantidad	Porcentaje	Tasa Of-Rate	Cantidad	Porcentaje
0	67.854	82,3%	60	101	0,1%
10	5.685	6,9%	70	226	0,3%
20	2.152	2,6%	80	61	0,1%
30	2.323	2,8%	90	2	0,0%
40	336	0,4%	100	2.143	2,6%
50	1.563	1,9%	TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

Tasa Win-Rate: Corresponde tasa de adjudicaciones sobre las ofertas enviadas. Es una medida del éxito de las ofertas presentadas por el proveedor. El indicador ha sido aproximado a las decenas por simplicidad.

Tabla 25: Variable: Tasa Win-Rate

Tasa Win-Rate	Cantidad	Porcentaje	Tasa Win-Rate	Cantidad	Porcentaje
0	41.470	50,3%	60	1.534	1,9%
10	3.979	4,8%	70	2.131	2,6%
20	4.082	5,0%	80	1.627	2,0%
30	5.403	6,6%	90	454	0,6%
40	2.529	3,1%	100	13.887	16,8%
50	5.350	6,5%	TOTAL	82.446	100,0%

(Fuente: Elaboración Propia)

5. CAPITULO 5: MODELADO

Al estar trabajando con una base que supera los 80.000 registros se vuelve necesario segmentarlos para facilitar la comprensión de los fenómenos. Como el foco del estudio es el comportamiento de fuga, se pondrá especial énfasis en los segmentos que contemplan este comportamiento.

La primera segmentación que se realizará es en base a tamaño de empresa. Se analizarán por separado las empresas según su tamaño. Principalmente debido a que si el estudio contempla el análisis de la información transaccional de los proveedores para caracterizar su comportamiento, comprender que el volumen de transacciones que mueven las microempresas versus las grandes empresas es fundamental. Al hacer esta división se estima que la varianza de ciertas variables disminuya (por ejemplo de los montos transados, ofertas enviadas y órdenes de compra recibidas). .

Tabla 26: Segmentación Tamaño Empresa según ventas anuales

Tamaño Empresa	Ventas Anuales en UF
Micro	Menos de 2.400
Pequeña	2.401 a 25.000
Mediana	25.000 a 100.000
Grande	Sobre 100.000

(Fuente: "Observatorio Empresas" (Observatorio Empresas, 2010))

5.1. **Aplicación de Técnicas de Clusterización:**

Se cuenta con una base de datos por tamaño de empresa, con 14 variables que describen a cada instancia de proveedor (tabla 7). Se busca encontrar una segmentación para cada conjunto a través de la aplicación de técnicas de clusterización. En esta parte, se aplicará la técnica de clusterización empleando las 14 variables descritas en el capítulo anterior.

El algoritmo definido para realizar la clusterización es el algoritmo de k-medias, el cual entregara los k segmentos de proveedores, según un k que recibe como parámetro.

Para encontrar ese k óptimo por segmento se aplicará otro algoritmo de clusterización, llamado Algoritmo EM (Expectation Maximization). El principal uso de este algoritmo es que a diferencia de k-medias, no requiere obligatoriamente un determinado número de clusters, sino que el mismo algoritmo entrega el cluster óptimo, buscando los clusters más probables dados los datos.

El resultado de la aplicación de EM, es el siguiente:

Tabla 27: Número de Clusters por segmento empresa

Tamaño Empresa	Cantidad de Registros	Clusters Según EM
MICRO	59.140	5
PEQUEÑA	16.548	5
MEDIANA	4.344	5 ²⁰ (7)
GRANDE	2.414	4
TOTAL	82.446	6²¹

(Fuente: Elaboración Propia)

Posteriormente, se aplica el algoritmo k-medias obteniendo los clusters para su posterior interpretación.

Si bien con el algoritmo EM se obtiene un número de clusters basado en la información empleada para el entrenamiento, este número está sujeto a interpretación y evaluación según reglas de negocio. Se considera que lo más importante desde aquí, más allá de la técnica es la interpretación según reglas de negocio.

Es importante mencionar que para facilitar la comprensión de los clusters, se utilizarán etiquetas. Estas etiquetas intentarán explicar intuitivamente cada cluster, y se utilizarán las mismas siempre, aunque por tamaño de empresa difieran en variables transacciones (en términos de volumen).

Las etiquetas utilizadas son las siguientes (se presenta en orden de transaccionalidad, desde el que menos transa al más intenso en transacciones):

- Proveedores “Fugados”
- Proveedores “Candidatos a Fuga”
- Proveedores “Nuevos”
- Proveedores “Esporádicos u ocasionales”
- Proveedores “Alto Potencial”
- Proveedores “Leales”

²⁰ El algoritmo EM arroja que 7 es el número óptimo de clústeres en el segmento MEDIANO. Por simplicidad se evaluarán 5 clústeres. Al realizar la evaluación con 7, es posible encontrar clústeres que se diferencian por sutilezas que sin pérdida de generalidad se pueden agrupar.

²¹ Se ha aplicado también el test a la totalidad de la base a modos exploratorio.

5.1.1. Empresas Micro

En las empresas clasificadas como micro se obtienen 5 segmentos.

Tabla 28: Clusters: Empresas Micro

<i>VARIABLE</i>	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>	<i>Cluster 4</i>	<i>Cluster 5</i>
<i>Etiqueta</i>	<i>Fugados</i>	<i>Candidatos a fuga</i>	<i>Nuevos</i>	<i>Alto potencial</i>	<i>Leales</i>
<i>Cantidad Proveedores</i>	5.816	12.892	11.657	13.494	15.281
<i>% proveedores</i>	9,8%	21,8%	19,7%	22,8%	25,8%
<i>Estado Transaccional</i>	Fugado	Transando	Transando	Transando	Transando
<i>Año</i>	[2003 a 2006]	[2009]	[2009]	[2009]	[2003 a 2006]
<i>Registrado en ChileProveedores</i>	SI	NO	NO	NO	SI
<i>Vigente en ChileProveedores</i>	NO	NO	NO	NO	SI
<i>Reclamos</i>	NO	NO	NO	NO	NO
<i>Cantidad de Meses Logeado</i>	5,4 (±3,3)	1,7 (±2,5)	3,4 (±2,6)	3,6 (±2,7)	9,0 (±3,2)
<i>of-Rate</i>	9,0 (±24,3)	0,9 (±6,1)	14,6 (±31,3)	2,8 (±13,4)	5,1 (±11)
<i>Win-Rate</i>	5,2 (±20,7)	3,1 (±10,8)	0,07 (±1,4)	73,4 (±37,7)	45,2 (±32,3)
<i>Cantidad Adjudicaciones</i>	[0]	[1]	[0]	[1]	[3 a 10]
<i>Cantidad Órdenes de Compra</i>	[0]	[1]	[0]	[1]	[2 a 4]
<i>Cantidad Ofertas</i>	[1]	[0]	[1]	[1]	[5 a 15]
<i>Monto Transado (USD)</i>	[0]	[1 a 1m]	[0]	[1m a 5m]	[5m a 50m]
<i>Ofertas Rechazadas</i>	[0]	[0]	[0]	[0]	[0]
<i>Promedio Logeo (mensual)</i>	[1 a 5]	[0]	[1 a 5]	[1 a 5]	[6 a 15]

(Fuente: Elaboración Propia)

Descripción de los Clústeres:

Clúster 1 – Proveedores Fugados (9,8%): Segmento de proveedores caracterizado por tener en su mayoría proveedores en estado fugado, que se registraron alguna vez

en ChileProveedores, pero no se mantienen vigentes. Poseen una tasa de éxito baja y una tasa de ofertas rechazadas alta, lo que indica que son proveedores no exitosos y eventualmente desencantados del sistema.

Clúster 2 – Proveedores Candidatos a fuga (21,8%): Proveedores de comportamiento de transacción muy similar a los proveedores nuevos del clúster 3. Tienen una tasa de éxito y nivel de transaccional bajo. Como su logeo es muy bajo, se estima que son proveedores ocasionales, que eventualmente pueden desertar del sistema.

Clúster 3 – Proveedores Nuevos (19,7%): Proveedores principalmente nuevos (77%), con una muy baja tasa de éxito, nivel de envío de ofertas (por ende bajos montos de adjudicación). También presentan una alta tasa de ofertas rechazadas y un alto nivel de logeo (lo que se puede interpretar como que son proveedores en etapa de aprendizaje del sistema). Agrupa un sector no menor de proveedores fugados, Si bien, estos no son nuevos, su comportamiento transaccional es similar a uno nuevo.

Clúster 4 – Proveedores Alto Potencial (22,8%): Proveedores también principalmente nuevos (desde 2009), con una alta tasa de éxito (promediando sobre el 70%), pero que transan por montos más bajos (en general volumen transaccional menor). Se caracterizan por que en su mayoría no están registrados en ChileProveedores. En el caso del sector micro, son proveedores que participan en procesos puntuales.

Clúster 5 - Proveedores Leales (25,8%): Segmento de proveedores caracterizado por un alto nivel de transaccionalidad (95%), registro en ChileProveedores (y vigencia). Son intensos también en meses de participación, logeo mensual. Tienen una tasa de éxito relativamente alta (promediando un 45%) Poseen un alto nivel de oferta y recepción de Órdenes de Compra.

5.1.2. Empresas Pequeñas

En las empresas clasificadas como pequeñas se obtienen 5 segmentos.

Tabla 29: Clusters: Empresas Pequeñas

VARIABLE	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Nombre	Candidatos a Fuga	Nuevos	Esporádicos	Alto Potencial	Leales
Cantidad Proveedores	4.351	3.284	2.378	3.453	3.082
% proveedores	26,3%	19,8%	14,4%	20,9%	18,6%
<i>Estado Transaccional</i>	Transando	Transando	Transando	Transando	Transando

Año	[2003 a 2006]	[2009]	[2003 a 2006]	[2003 a 2006]	[2003 a 2006]
Registrado en ChileProveedores	NO	NO	SI	SI	SI
Vigente en ChileProveedores	NO	NO	SI	SI	SI
Reclamos	NO	NO	NO	NO	NO
Cantidad de Meses Logeado	2,2 ($\pm 2,8$)	4,7 ($\pm 3,1$)	5,7 (± 3)	9,6 ($\pm 2,6$)	11,4 ($\pm 1,4$)
of-Rate	1,9 ($\pm 11,1$)	15,4 ($\pm 30,9$)	5,8 ($\pm 16,0$)	6,0 ($\pm 10,8$)	4,3 ($\pm 7,6$)
Win-Rate	19,6 (± 37)	2,9 ($\pm 15,6$)	55,8 ($\pm 38,5$)	47,4 ($\pm 30,6$)	31,8 ($\pm 24,4$)
Cantidad Adjudicaciones	[1]	[0]	[1]	[3 a 10]	[>10]
Cantidad Órdenes de Compra	[1]	[0]	[1]	[5 a 15]	[>15]
Cantidad Ofertas	[0]	[1]	[2 a 4]	[5 a 15]	[>15]
Monto Transado (USD)	[1 a 1m]	[0]	[1m a 5m]	[5m a 50m]	[50m a 500m]
Ofertas Rechazadas	[0]	[0]	[0]	[0]	[0]
Promedio Logeo (mensual)	[1 a 5]	[6 a 15]	[1 a 5]	[6 a 15]	[16 a 50]

(Fuente: Elaboración Propia)

Descripción de los Clusters:

Clúster 1 – Proveedores Candidatos a fuga (26,3%): Proveedores que se caracterizan por estar en el sistema de antes de 2009, que no interactúan en ChileProveedores. Se logean pocos meses en el año, presentan adjudicaciones y una tasa de éxito regular. Principalmente transan en montos bajos. Concentra un importante porcentaje de proveedores fugados.

Clúster 2 – Proveedores nuevos (19,8%): Segmento de proveedores en su mayoría activos pero con intensidad en fugados (41%). Es el segmento de proveedores que agrupa a aquellos con bajo nivel de presencia mensual en plataforma, registro en ChileProveedores, alta tasa de ofertas rechazadas y win-rate muy bajo. Niveles de transacción bajísimos y se caracteriza por ser principalmente proveedores nuevos (2009). Es un grupo que comparte características con los Candidatos a fuga de la microempresa.

Clúster 3 - Proveedores Esporádicos (14,4%): Segmento de proveedores con comportamiento muy similar a los proveedores de alto potencial (pero menor). La diferencia está en que ese grupo interactúa con menor frecuencia y volumen que el anterior, manteniendo el nivel de éxito en el sistema (presentan la mayor tasa de éxito del conjunto, lo que habla de un gran nivel de efectividad en sus transacciones).

Clúster 4 - Proveedores Alto Potencial (20,9%): Segmento de proveedores caracterizado por un alto nivel de actividad (sobre el 47%), registro en ChileProveedores (y vigencia). Son intensos también en meses de participación. Poseen un alto nivel de oferta, tasa de éxito y recepción de Órdenes de Compra.

Clúster 5 – Proveedores leales (18,6%): Segmento de proveedores que presenta mayor intensidad en transacciones con relación a los otros. Son proveedores activos, antiguos en su mayoría, registrados y vigentes en ChileProveedores. Presentes en casi todos los meses del año, con alta tasa de éxito, altos niveles de transacción e interacción mensual en plataforma.

5.1.3. Empresas Medianas

En las empresas clasificadas como medianas se obtienen 5 segmentos.

Tabla 30: Clusters: Empresas Medianas

VARIABLE	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Nombre	Fugados	Candidatos a fuga	Esporádicos	Alto Potencial	Leales
Cantidad Proveedores	702	942	1.222	599	879
% proveedores	16,2%	21,7%	28,1%	13,8%	20,2%
<i>Estado Transaccional</i>	Fugado	Transando	Transando	Transando	Transando
<i>Año</i>	[2003 a 2006]	[2003 a 2006]	[2003 a 2006]	[2003 a 2006]	[2003 a 2006]
<i>Registrado en ChileProveedores</i>	NO	NO	SI	SI	SI
<i>Vigente en ChileProveedores</i>	NO	NO	SI	SI	SI
<i>Reclamos</i>	NO	NO	NO	NO	NO
<i>Cantidad de Meses Logeado</i>	1,3 (±2,4)	4,2 (±3,0)	7,9 (±3,2)	10,4 (±2,4)	11,7 (±1,1)
<i>of-Rate</i>	2,1 (±12,8)	7,6 (±22,8)	5,7 (±16,5)	14,7 (±20,4)	3,9 (±7,1)
<i>Win-Rate</i>	3,6 (±17,7)	28,4 (±41,0)	42,7 (±36,6)	26,0 (±24,9)	37,1 (±22,6)

<i>Cantidad Adjudicaciones</i>	[1]	[1]	[3 a 10]	[3 a 10]	[>10]
<i>Cantidad Órdenes de Compra</i>	[1]	[1]	[5 a 15]	[>15]	[>15]
<i>Cantidad Ofertas</i>	[0]	[0]	[2 a 4]	[5 a 15]	[>15]
<i>Monto Transado (USD)</i>	[1 a 1m]	[1 a 1m]	[5m a 50m]	[50m a 500m]	[50m a 500m]
<i>Ofertas Rechazadas</i>	[0]	[0]	[0]	[2 a 4]	[5 a 15]
<i>Promedio Logeo (mensual)</i>	[0]	[1 a 5]	[6 a 15]	[16 a 50]	[16 a 50]

(Fuente: Elaboración Propia)

Descripción de los Clusters:

Clúster 1 – Proveedores Fugados (16,2%): Segmento de proveedores caracterizado por tener en su mayoría proveedores en estado fugado. Algunos se registraron alguna vez en ChileProveedores, pero no se mantienen vigentes. Poseen una tasa logeos y de éxito bajísima, lo que indica que son proveedores no exitosos y eventualmente desencantados del sistema.

Clúster 2 – Proveedores Candidatos a fuga (21,7%): Proveedores que se caracterizan por estar en el sistema de antes de 2009, que no interactúan en ChileProveedores. Se logean pocos meses en el año, presentan adjudicaciones y una tasa de éxito regular. Principalmente transan en montos bajos.

Clúster 3 - Proveedores Esporádicos (28,1%): Segmento de proveedores con comportamiento muy similar a los proveedores con alto potencial. La diferencia está en que ese grupo interactúa con menor frecuencia y volumen que ellos, manteniendo el nivel de éxito en el sistema (presentan la mayor tasa de éxito del conjunto, lo que habla de un gran nivel de efectividad en sus transacciones).

Clúster 4 - Proveedores Exitosos (13,8%): Segmento de proveedores caracterizado por un alto nivel de actividad, registro en ChileProveedores (y vigencia). Son intensos también en meses de participación. Poseen un alto nivel de oferta, tasa de éxito y recepción de Órdenes de Compra. Se diferencia de los leales por tener una menor intensidad de transacciones.

Clúster 5 – Proveedores leales (20,2%): Segmento de proveedores que presenta alta intensidad en transacciones. Son proveedores activos, antiguos en su mayoría, registrados y vigentes en ChileProveedores. Presentes en casi todos los meses del año, con alta tasa de éxito, altos niveles de transacción e interacción mensual en plataforma.

5.1.4. Empresas Grandes

En las empresas clasificadas como grandes se obtienen 4 segmentos.

Tabla 31: Clusters: Empresas Grandes

VARIABLE	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Nombre	Candidatos a fuga	Esporádicos	Alto potencial	Leales
Cantidad Proveedores	339	663	559	853
% proveedores	14,0%	27,5%	23,2%	35,3%
<i>Estado Transaccional</i>	Transando	Transando	Transando	Transando
<i>Año</i>	[2009]	[2003 a 2006]	[2003 a 2006]	[2003 a 2006]
<i>Registrado en Chile Proveedores</i>	SI	NO	SI	SI
<i>Vigente en Chile Proveedores</i>	SI	NO	SI	SI
<i>Reclamos</i>	NO	NO	NO	NO
<i>Cantidad de Meses Logeado</i>	6,7 (±3,6)	2,5 (±3,0)	9,4 (±2,8)	11,7 (±0,9)
<i>of-Rate</i>	15,1 (±28,9)	2,1 (±11,7)	6,3 (±13,0)	4,7 (±8,5)
<i>Win-Rate</i>	7,9 (±22,1)	18,1 (±35,8)	55,2 (±31,9)	38,9 (±21,7)
<i>Cantidad Adjudicaciones</i>	[0]	[1]	[3 a 10]	[>10]
<i>Cantidad Órdenes de Compra</i>	[0]	[1]	[5 a 15]	[>15]
<i>Cantidad Ofertas</i>	[2 a 4]	[0]	[5 a 15]	[>15]
<i>Monto Transado</i>	[0]	[1 a 1m USD]	[5m a 50m USD]	[50m a 500m USD]
<i>Ofertas Rechazadas</i>	[0]	[0]	[>15]	[5 a 15]
<i>Promedio Logeo (mensual)</i>	[6 a 15]	[1 a 5]	[6 a 15]	[16 a 50]

(Fuente: Elaboración Propia)

Descripción de los Clusters:

Clúster 1 - Proveedores Candidatos a Fuga (14%): Segmento que posee una mayor cantidad de proveedores activos (69%) y ser en su mayoría nuevos proveedores (desde 2009). Posee tasas de ofertas rechazadas alta, y una tasa de éxito baja, bajo nivel de adjudicación y de logeo.

Clúster 2 - Proveedores Esporádicos (27,5%): Segmento de proveedores que interactúa con menor frecuencia y volumen que los demás. Poseen una tasa de éxito regular-baja y no utilizan en general los servicios como ChileProveedores. Poseen un porcentaje no menor de proveedores en estado fugado.

Clúster 3 - Proveedores Alto Potencial (23,2%): Segmento de proveedores caracterizado por un alto nivel de actividad, registro en ChileProveedores (y vigencia). Son intensos también en meses de participación. Poseen un alto nivel de oferta, tasa de éxito y recepción de Órdenes de Compra. Se diferencia de los leales por tener una menor intensidad de transacciones.

Clúster 4 – Proveedores leales (35,3%): Segmento que agrupa a los proveedores más intensos en oferta, adjudicación, montos transados y uso en general de los servicios de ChileCompra. Poseen una tasa de éxito del 38% promedio. Y corresponde principalmente a proveedores antiguos. Es el segmento que agrupa los montos más elevados de todos los proveedores analizados.

5.1.5. Comentarios generales sobre la clusterización

Se cuenta con una segmentación en clusters para cada conjunto de empresas. El paso a seguir corresponde a evaluar la clasificación con miras a predecir en qué categoría de clasificación se puede ubicar una empresa proveedora en el futuro.

Se ha optado por emplear las mismas leyendas para los grupos cluster-tamaño empresa para facilitar la comprensión. En algunos segmentos de empresas por tamaño no se encuentran algunos tipos (por ejemplo en las medianas y grandes no hay clusters del tipo nuevo), debido a que sus características no resaltan a la luz de los datos. Por ejemplo, en las grandes empresas no es significativa la presencia de proveedores fugados como agrupación, esto tiene sentido desde la mirada de que existen pocas empresas grandes en esta clasificación. En el caso de las pequeñas, tampoco existe dicho cluster, pero si existen proveedores fugados. El problema es que no son homogéneos y se encuentran diluidos en las otras aglomeraciones.

Tabla 32: Resumen Clusters

	Micro	Pequeña	Mediana	Grande	Total
Fugado	5.816		702		6.518
	9,8%		16,2%		7,9%
Candidato a fuga	12.892	4351	942	339	18.524
	21,8%	26,3%	21,7%	14,0%	22,5%
Nuevo	11.657	3284			14.941
	19,7%	19,8%			18,1%
Esporádico		2378	1222	663	4.263
		14,4%	28,1%	27,5%	5,2%
Alto potencial	13.494	3453	599	559	18.105
	22,8%	20,9%	13,8%	23,2%	22,0%
Leal	15.281	3082	879	853	20.095
	25,8%	18,6%	20,2%	35,3%	24,4%
TOTAL	59.140	16.548	4.344	2.414	82.446

(Fuente: Elaboración Propia)

En la tabla anterior se muestra un resumen de los clusters encontrados por segmentos de empresa. Esta clasificación fue validada por el área de Desarrollo Empresarial y será considerada en las próximas estrategias para referirse a los tipos de proveedor y eventuales segmentos objetivos.

Los grupos “fugados” y “candidatos a fuga” resultan ser los grupos más sensibles en términos de satisfacción y estado transaccional (agrupan a la mayoría de los proveedores en estado fugado o a los que se le “parecen” en términos de actividad transaccional). Si un proveedor es clasificado en alguna de estas categorías tiene una alta probabilidad de fugarse.

Por otro lado, los proveedores en los clusters “esporádicos”, “alto potencial” y “leales” corresponden a los proveedores intensos en transaccionalidad y uso de los servicios de la Dirección de Compra.

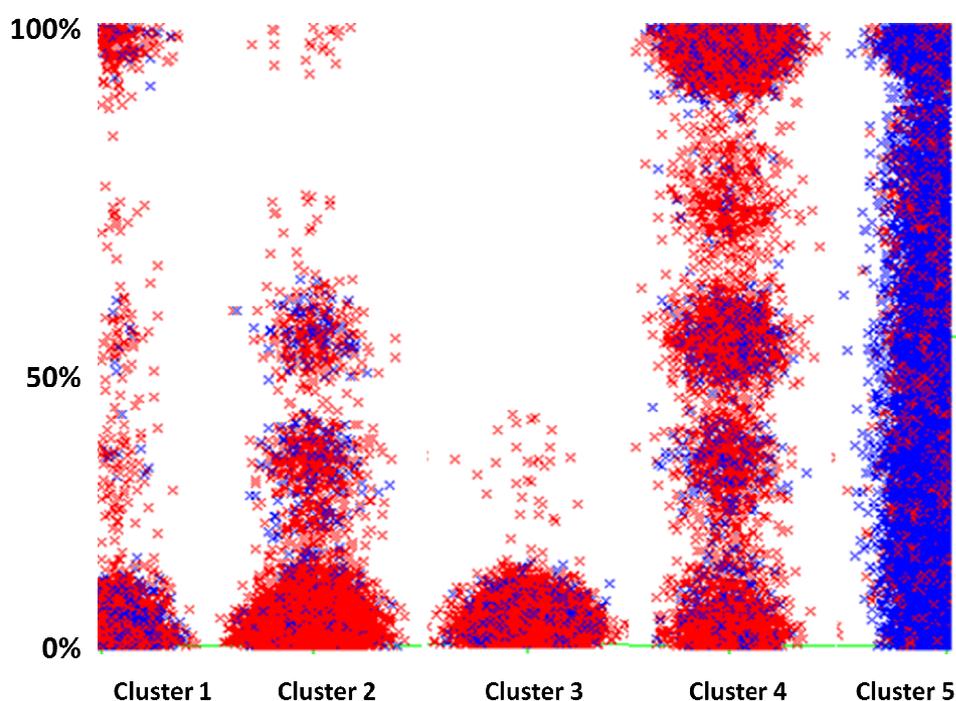
5.1.6. Gráficos de dispersión de clústeres

Para ilustrar el impacto de la clasificación a continuación se exponen algunas ilustraciones con gráficos de dispersión, en donde se observará como ciertas variables discriminan y permiten segmentar.

En la microempresa se puede observar que el Cluster 5 (leales), es el único intenso en vigencia en ChileProveedores, con una concentración del 80%.

Si se analiza el registro, la concentración aumenta al 95% de los proveedores del cluster con esa condición. El cluster 4 agrupa a quienes poseen tasa win-rate intermedias a altas (proveedores de alto potencial), con tasas que oscilan cerca del 73% de éxito. El cluster 3 (nuevos) se observa que posee una tasa de éxito muy baja y una tasa de falla (of-rate) muy alta en comparación a los otros segmentos encontrados. Por último, se observa que los candidatos a fuga (cluster 2) poseen win rate bajo el 50%, lo que representa un candidato a predictor. En la siguiente ilustración se podrán observar con mayor detalle estas caracterizaciones.

Ilustración 31: Cluster v/s Win-Rate (Microempresas)



Eje X: Clusters, Eje Y: Win-Rate

Color Azul: Vigente en ChileProveedores, Color Rojo: No Vigente en ChileProveedores

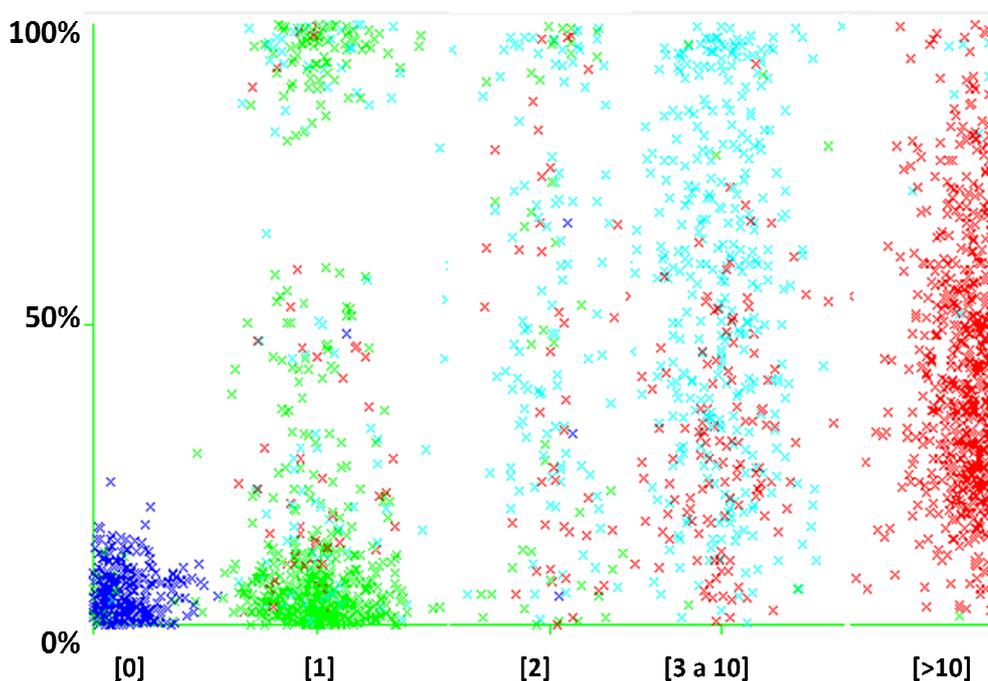
(Fuente: Weka Clusterer Visualize)

Otro ejemplo de dispersión se puede observar en las grandes empresas al contrastar el nivel de adjudicaciones con la tasa de éxito y la segmentación encontrada. Se puede observar en la ilustración que el Cluster 1 (Candidatos a fuga), está centrado en win-rate 7,9% y 0 adjudicaciones, lo que representa un bajo nivel de actividad (Color AZUL). Por otro lado, el Cluster 2 (Esporádicos), se centra en una tasa win-rate 18% y 1 adjudicación (Color VERDE).

Y ahora se puede contrastar con los clusters de mayor potencial de oferta, como por ejemplo el Cluster 3 (Alto potencial), el cual está centrado en win-rate 55% y desde 2 a 10 adjudicaciones (Color CELESTE). Finalmente el cluster mas intenso en adjudicaciones es el Cluster 4 (Leales), Centrado en win-rate 38,9% y sobre 10 adjudicaciones (Color ROJO).

Si en vez de cruzar con adjudicaciones, la comparación se realiza con la adjudicación de órdenes de compra la dispersión es muy similar, pero los volúmenes aumentan, pues las grandes empresas tienen alta participación en los Convenios Marco de la institución los que generan alta intensidad de Órdenes de Compra.

Ilustración 32: Cantidad de Adjudicaciones v/s WinRate (Grandes Empresas)



Eje X: Cantidad Adjudicaciones, Eje Y: Win-Rate

Colores: CLUSTERS

(Fuente: Weka Clusterer Visualize)

5.2. Aplicación de Técnicas de Selección de Atributos

Previo a la clasificación, se realizará una etapa de preselección de los atributos que serán utilizados. Utilizar todas las variables se vuelve extremadamente complejo a nivel de análisis y de interpretación, por ende se utilizarán filtros de atributos que el mismo software posee. La idea es clasificar a las variables en función a su peso en la información. Si una variable no agrega valor al análisis ya sea por estar correlacionada con otra o por no tener la calidad suficiente para representar clusters, serán eliminadas del análisis.

Anteriormente en la clusterización se emplearon las 14 variables, si bien, en la interpretación de los clusters se observa que existen algunas variables que no distinguen entre clusters, o sea, que eventualmente podrían descartarse, sin perder generalidad de la clasificación. Casos como este corresponden a la cantidad de reclamos, el registro en ChileProveedores (que está implícito en la vigencia en el registro), etc.

Para realizar este filtrado, se utilizarán tres métodos de selección de atributos para cada uno de los segmentos de empresa, incluyendo todas las variables. Estos métodos originan un ranking según su criterio. Al realizar las comparaciones, es necesario indicar un atributo como “clase”, es decir, con respecto a que variable se evaluará el aporte al modelo. En este caso se evaluará con respecto a la clasificación realizada anteriormente (clusters). Los métodos son los siguientes:

- InfoGainAttributeEval: Evalúa los atributos midiendo la ganancia de información de cada uno con respecto a la clase.
- GainRatioAttributeEval: Evalúa cada atributo midiendo su razón de beneficio con respecto a la clase.
- ChiSquaredAttributeEval: Calcula el valor estadístico Chi-cuadrado de cada atributo con respecto a la clase y así obtiene el nivel de correlación entre la clase y cada atributo

En la siguiente tabla se encuentran las variables que por los 3 métodos han sido seleccionadas y las variables que se descartarán. En todos los segmentos de empresa los resultados son similares. En el anexo F, se detallan los resultados por método para cada segmento de empresa.

Tabla 33: Variables seleccionadas para Clasificación

Variables Seleccionadas	Variables No Seleccionadas
Cantidad Órdenes de Compra	Cantidad Ofertas rechazadas
Estado Transaccional	Promedio Logeo
Monto Transado	Registrado en ChileProveedores
Cantidad Adjudicaciones	Tasa Of-Rate
Cantidad Ofertas	Año Primera interacción
Tasa: Win rate	Reclamos 2009
Cantidad de meses logeado	
Vigente en ChileProveedores	

(Fuente: Elaboración Propia)

La variable Cantidad de ofertas rechazadas y la Tasa Of-rate tiene sentido descartarlas pues su intensidad es muy baja, no genera grandes diferencias pues se acumula en 0. El registro en ChileProveedores está implícito en la variable Vigencia en ChileProveedores. El nivel de reclamos no diferencia entre clusters. Si bien el estado transaccional no es considerado con aporte significativo en los modelos, se considerará igual, por ser parte del estudio.

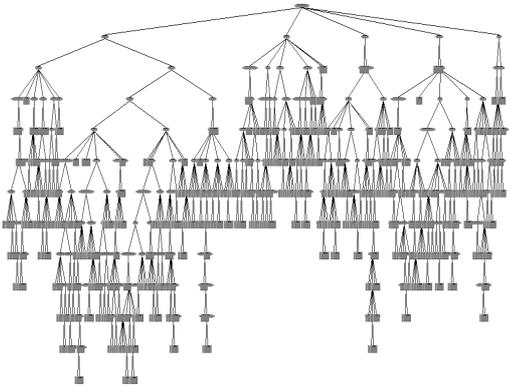
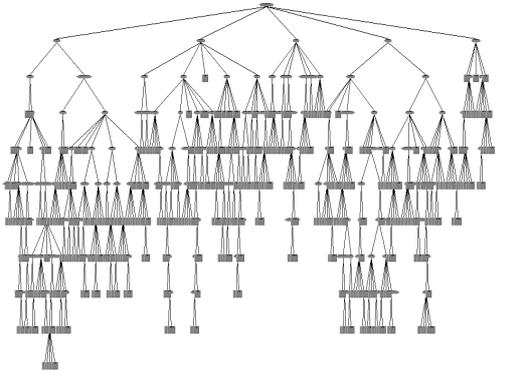
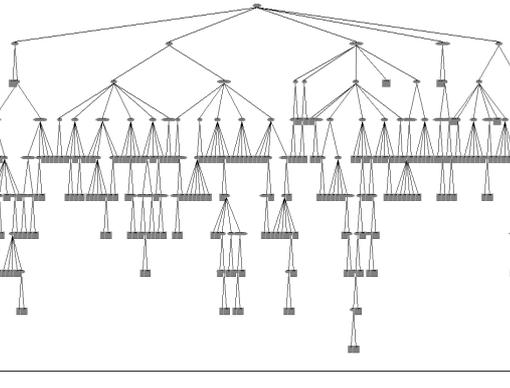
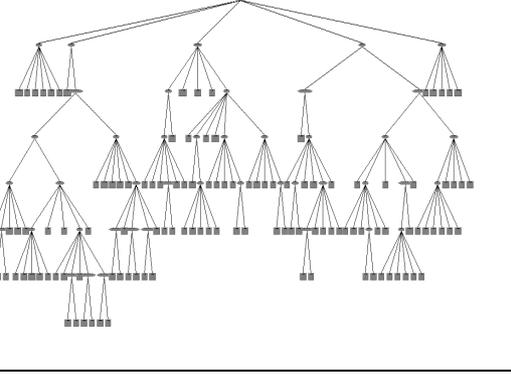
Estas observaciones están alineadas con la interpretación según criterios de negocio, lo que genera un nuevo set de 8 variables, las cuales se emplearán en el siguiente paso del análisis.

5.3. Aplicación de Técnicas de Clasificación

5.3.1. Árboles de Decisión

Finalmente, el resultado de la clasificación arrojó árboles de decisión para cada categorización de proveedores. La variable objetivo en este caso corresponde al atributo cluster. Los árboles tienen la forma que se observa en las siguientes ilustraciones:

Ilustración 33: Árboles de Decisión por Tamaño de Empresa

Microempresa	Empresas Pequeñas
	
Tamaño del árbol: 764 elementos Número de Hojas: 493	Tamaño del árbol: 564 elementos Número de Hojas: 382
Empresas Medianas	Empresas Grandes
	
Tamaño del árbol: 305 elementos Número de Hojas: 199	Tamaño del árbol: 164 elementos Número de Hojas: 120

(Fuente: Weka Classifier Tree Visualizer)

Se observa que a medida que aumenta el tamaño de la empresa los árboles se van descomplejizando. Esto se debe a que la cantidad de proveedores va disminuyendo

considerablemente y se van encontrando con mayor seguridad conjuntos de empresas homogéneos.

En particular el sector micro es muy complejo de analizar, la cantidad de proveedores y las diversas configuraciones de variables que existen generan que la complejidad de la interpretación del árbol sea mayor.

El árbol de decisión permite clasificar proveedores que no estén incluidos en el modelo en una categoría particular. Se dirá entonces que el conjunto empleado hasta el momento es el conjunto de “entrenamiento”.

Es posible utilizar un nuevo conjunto de proveedores o el mismo, pero con variables considerando otro intervalo de tiempo (por ejemplo, desde junio 2009 a junio 2010). De esta forma para el nuevo conjunto de proveedores se obtendrá una distribución de probabilidad de ubicación en cada cluster, es decir, se intentará predecir dónde debiese ser clasificado.

En la siguiente ilustración se muestra un conjunto de datos no clasificado. El cual será evaluado utilizando el modelo de árboles de decisión diseñado.

Ilustración 34: Ejemplo de Datos no Clasificados

Estado Nominal	Vigente CHP Nominal	Cantidad de meses logeado Numeric	Win rate 2009 Numeric	Cantidad Adjudicaciones Nominal	Cantidad OC Nominal	Cantidad Ofertas Nominal	Monto USD Nominal	CLASE Nominal
Fugado	SI	5.0	0.0	[0]	[0]	[1]	[0]	
Trans...	NO	0.0	0.0	[1]	[>15]	[0]	[50000 a ...	
Trans...	NO	2.0	100.0	[1]	[1]	[1]	[50000 a ...	
Trans...	SI	10.0	100.0	[2]	[2 a 4]	[2 a 4]	[>500000...	
Trans...	SI	12.0	50.0	[3 a 10]	[>15]	[>15]	[>500000...	
Trans...	SI	12.0	30.0	[3 a 10]	[>15]	[>15]	[>500000...	
Trans...	NO	12.0	50.0	[3 a 10]	[5 a 15]	[5 a 15]	[>500000...	
Trans...	SI	12.0	10.0	[3 a 10]	[5 a 15]	[>15]	[50000 a ...	
Trans...	SI	12.0	10.0	[>10]	[>15]	[>15]	[>500000...	
Trans...	SI	12.0	0.0	[0]	[0]	[5 a 15]	[0]	
Trans...	SI	7.0	40.0	[2]	[2 a 4]	[5 a 15]	[1000 a 5...	
Trans...	NO	1.0	0.0	[1]	[>15]	[0]	[1000 a 5...	
Fugado	NO	6.0	0.0	[0]	[0]	[1]	[0]	
Trans...	NO	0.0	0.0	[1]	[2 a 4]	[0]	[1 a 1000 ...	
Trans...	SI	11.0	60.0	[3 a 10]	[5 a 15]	[5 a 15]	[50000 a ...	
Fugado	NO	0.0	0.0	[1]	[2 a 4]	[0]	[1 a 1000 ...	
Trans...	SI	5.0	100.0	[3 a 10]	[5 a 15]	[2 a 4]	[5000 a 5...	
Trans...	SI	10.0	0.0	[0]	[0]	[2 a 4]	[0]	
Trans...	SI	12.0	50.0	[3 a 10]	[>15]	[5 a 15]	[50000 a ...	
Trans...	SI	12.0	50.0	[3 a 10]	[>15]	[5 a 15]	[50000 a ...	
Trans...	NO	11.0	0.0	[0]	[0]	[5 a 15]	[0]	
Trans...	SI	10.0	70.0	[>10]	[>15]	[>15]	[50000 a ...	
Fugado	NO	4.0	0.0	[0]	[0]	[1]	[0]	

(Fuente: Weka Data Viewer)

La aplicación del modelo de árbol da el siguiente resultado predictivo, el cual es consistente con las descripciones de los clusters expresado en el capítulo anterior.

La tabla siguiente contiene a los 24 proveedores de la ilustración anterior (ejemplo) y la distribución de probabilidad que determina el cluster más probable en el que deba ser clasificado cada caso.

Tabla 34: Cluster Probable y Distribución de Probabilidades

PROVEEDOR	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Predicción
Caso 1	0.952	0	0.048	0	Cluster Candidato Fuga
Caso 2	0.004	0.006	0.968	0.022	Cluster Esporádico
Caso 3	0.004	0.006	0.968	0.022	Cluster Esporádico
Caso 4	0.314	0.029	0	0.657	Cluster Alto Potencial
Caso 5	0	0.955	0	0.045	Cluster Leales
Caso 6	0	0.955	0	0.045	Cluster Leales
Caso 7	0	0.079	0.008	0.913	Cluster Alto Potencial
Caso 8	0	0.956	0	0.044	Cluster Leales
Caso 9	0	0.992	0	0.008	Cluster Leales
Caso 10	0.952	0	0.048	0	Cluster Candidato Fuga
Caso 11	0.314	0.029	0	0.657	Cluster Alto Potencial
Caso 12	0.004	0.006	0.968	0.022	Cluster Esporádico
Caso 13	0.952	0	0.048	0	Cluster Candidato Fuga
Caso 14	0.004	0.006	0.968	0.022	Cluster Esporádico
Caso 15	0	0.079	0.008	0.913	Cluster Alto Potencial
Caso 16	0.004	0.006	0.968	0.022	Cluster Esporádico
Caso 17	0	0.031	0	0.969	Cluster Alto Potencial
Caso 18	0.952	0	0.048	0	Cluster Candidato Fuga
Caso 19	0	0.079	0.008	0.913	Cluster Alto Potencial
Caso 20	0	0.079	0.008	0.913	Cluster Alto Potencial
Caso 21	0.952	0	0.048	0	Cluster Candidato Fuga
Caso 22	0	0.992	0	0.008	Cluster Leales
Caso 23	0.952	0	0.048	0	Cluster Candidato Fuga
Caso 24	0.004	0.006	0.968	0.022	Cluster Esporádico

(Fuente: Elaboración Propia)

5.3.2. Reglas de Decisión

Las reglas de decisión obtenidas en cada caso permiten explicar y validar la composición de cada cluster formado. Al contrastar con los árboles de decisión, se observa que existe consistencia entre ambos modelos.

En la siguiente tabla se muestra cuántas reglas el modelo fue capaz de encontrar en cada tamaño de empresa clusterizado:

Tabla 35: Cantidad de Reglas por Tamaño de Empresa

	Micro	Pequeña	Mediana	Grande
Fugados	54		14	
Candidatos a Fuga	73	92	45	8
Nuevos	25	19		
Esporádicos		89	73	15
Alto Potencial	134	132	36	38
Leales	169	62	19	23
TOTAL	455	394	187	84

(Fuente: Elaboración Propia)

En el anexo I se encuentran las principales reglas por tamaño de empresa.

5.4. Evaluación de la Clasificación Realizada

En el capítulo 3, se definen algunos indicadores (medidas de exactitud) para evaluar el modelo de clasificación, entre los cuales se encuentran la Tasa TP, la tasa FP, la precisión, la Matriz de Confusión y el Coeficiente de Kappa.

Lo que sigue es una serie de indicadores que demuestran que la clasificación realizada es buena. La información se extrae principalmente de la matriz de confusión. Es importante destacar que una clasificación perfecta incluye la totalidad de los registros en la diagonal de la matriz.

5.4.1. Empresas Micro

- Instancias clasificadas correctamente: 52.135 (88,1%)
- Coeficiente de Kappa: 0,84

Tabla 36: Matriz de Confusión Micro

Micro	Alto Potencial	Nuevos	Fugados	Leales	Candidatos	Recall
Alto Potencial	12.678		19	701	96	94,0%
Nuevos	-	10.718	861	73	5	91,9%
Fugados	394	2.506	2.563	86	267	44,1%
Leales	901	47	61	14.043	229	91,9%
Candidatos	580	15	20	144	12.133	94,1%
PRECISIÓN	87,1%	80,7%	72,7%	93,3%	95,3%	

(Fuente: Elaboración Propia, Extraída de Weka Run Information)

5.4.2. Empresas Pequeñas

- Instancias clasificadas correctamente: 15.095 (91,2%)
- Coeficiente de Kappa: 0,889

Tabla 37: Matriz de Confusión Pequeña

Pequeña	Candidato	Leal	Nuevo	Alto Potencial	Esporádico	Recall
Candidato	4.257	4	3	53	34	97,8%
Leal	11	2.843	62	147	19	92,2%
Nuevo	91	15	3.121	28	29	95,0%
Alto Potencial	47	243	53	2.985	125	86,4%
Esporádico	213	31	86	159	1.889	79,4%
PRECISIÓN	92,2%	90,7%	93,9%	88,5%	90,1%	

(Fuente: Elaboración Propia, Extraída de Weka Run Information)

5.4.3. Empresas Medianas

- Instancias clasificadas correctamente: 3.792 (87,2%)
- Coeficiente de Kappa: 0,838

Tabla 38: Matriz de Confusión Mediana

Mediana	Esporádico	Candidato	Alto Potencial	Leal	Fugado	Recall
Esporádico	1.042	59	87	23	11	85,3%
Candidato	61	865	2	4	10	91,8%
Alto Potencial	122	28	426	18	5	71,1%
Leal	34	-	45	799	1	90,9%
Fugado	5	37	-	-	660	94,0%
PRECISIÓN	82,4%	87,5%	76,1%	94,7%	96,1%	

(Fuente: Elaboración Propia, Extraída de Weka Run Information)

5.4.4. Empresas Grandes

- Instancias clasificadas correctamente: 2.218 (91,8%)
- Coeficiente de Kappa: 0,888

Tabla 39: Matriz de Confusión Grande

Grande	Candidatos	Leales	Esporádicos	Alto Potencial	Recall
Candidatos	306	8	7	18	90,3%
Leales	5	786	5	56	92,3%
Esporádicos	20	1	627	16	94,4%
Alto Potencial	7	31	22	499	89,3%
PRECISIÓN	90,5%	95,2%	94,9%	84,7%	

(Fuente: Elaboración Propia, Extraída de Weka Run Information)

5.4.5. Comentarios generales sobre la evaluación

En general la clasificación posee una buena evaluación. Se cuenta con altos porcentajes de instancias clasificadas correctamente por el árbol de decisión (del orden del 90%).

El coeficiente de kappa resulta ser bastante bueno (dentro de los rangos teóricos que se consideran buenos, es decir, sobre 0,8).

Ahora, analizando lo que es el indicador “recall” o también conocido como la tasa “true-positive” (tasa de elementos que fueron clasificados en una categoría, con respecto a los que de verdad poseen esa categoría), los valores son buenos, ubicándose por lo general sobre el 85%.

Existe un caso especial y corresponde a un cluster en particular del sector micro. El cluster “fugados” posee un recall del 44%, esto significa que solo el 44% de los proveedores que debiesen ser clasificados en ese cluster, fueron clasificados allí. Existe un 43% de ellos que fueron clasificados por el árbol en ramas relacionadas con el cluster “nuevos”. La razón es que estos clusters son muy parecidos transaccionalmente. Difieren principalmente en el año de ingreso al sistema y su frecuencia de uso del mismo.

Finalmente analizando la precisión, es decir la proporción de elementos que efectivamente poseen la clase, entre todos los que fueron clasificados allí, se observa que los valores se mantienen en porcentajes sobre el 80% en general, lo que es considerado bueno.

A la luz de estos resultados se valida el uso del modelo como clasificador de instancias de proveedores. En el siguiente capítulo se realizará una clasificación de proveedores al año 2010 considerando las variables óptimas aquí descritas y una evaluación de los proveedores considerados en la base, con miras a evaluar si la predicción ha sido satisfactoria.

Al describir los datos se planteó la posibilidad de existencia de “ocultamiento de información” al delimitar el espacio temporal de los datos. Se adoptó el supuesto de que esta información no sería relevante en el análisis, lo que es validado con estos resultados.

6. CAPITULO 6: EVALUACIÓN Y APLICACIONES

A continuación se presenta la evaluación del modelo. Corresponde a la evaluación del estado transaccional de los proveedores analizados a junio de 2010, es decir 6 meses después de su última clasificación.

Esta evaluación intenta describir cual fue la dinámica transaccional de los proveedores después de ser clasificados.

Tabla 40: Evaluación al año 2010 (Base 2009) (Empresas Micro)

	Fugados		Candidatos a Fuga		Nuevos		Alto Potencial		Leales	
TOTAL	5.816		12.892		11.657		13.494		15.281	
<i>Estado Transaccional Diciembre 2009</i>										
Fugado	4.432	76,2%	4.201	32,6%	4.050	34,7%	3.338	24,7%	638	4,2%
Transando	1.384	23,8%	8.691	67,4%	7.607	65,3%	10.156	75,3%	14.643	95,8%
<i>Estado Transaccional Proveedores Transando 2009 a Junio 2010</i>										
Fugado	795	57,4%	4.998	57,5%	5.578	73,3%	5.315	52,3%	2.991	20,4%
Transando	589	42,6%	3.693	42,5%	2.029	26,7%	4.841	47,7%	11.652	79,6%
<i>Estado Transaccional Proveedores Fugados 2009 a Junio 2010</i>										
Fugado	3.474	78,4%	3.319	79,0%	3.639	89,9%	2.462	73,8%	387	60,7%
Transando	958	21,6%	882	21,0%	411	10,1%	876	26,2%	251	39,3%
<i>Estado Transaccional TODOS a Junio 2010</i>										
Fugado	4.269	73,4%	8.317	64,5%	9.217	79,1%	7.777	57,6%	3.378	22,1%
Transando	1.547	26,6%	4.575	35,5%	2.440	20,9%	5.717	42,4%	11.903	77,9%

(Fuente: Elaboración Propia)

La tabla 39 muestra los resultados a junio 2010 del estado transaccional de los proveedores del sector microempresarial que fueron clusterizados.

En general se puede extraer la siguiente información:

En el Clúster fugado: Quienes efectivamente estaban fugados en diciembre 2009 (76,2%), el 78% está fugado a junio de 2010. Quienes aun no estaban fugados en diciembre 2009 (23,8%), el 57,4% está fugado a junio de 2010.

En el Clúster Candidatos a fuga: Quienes efectivamente estaban fugados en diciembre 2009 (32,6%), el 79% está fugado a junio de 2010. Quienes aun no estaban fugados en diciembre 2009 (67,4%), el 57,5% está fugado a junio de 2010.

En el Clúster Nuevos: Quienes efectivamente estaban fugados en diciembre 2009 (34,7%), el 89,9% está fugado a junio de 2010. Quienes aun no estaban fugados en diciembre 2009 (65,3%), el 73,3% está fugado a junio de 2010.

Hasta ahora el modelo está siendo efectivo pues indica con alto nivel de certeza quiénes se fugarán. En particular se observa que el porcentaje más alto de fuga lo poseen los proveedores nuevos. Esto quiere decir que existen muchos proveedores del sector microempresarial que se registran a modo de “prueba del sistema”, pero que no logran satisfacer sus necesidades y lo abandonan.

En el Clúster de Alto Potencial: Quienes efectivamente estaban fugados en diciembre 2009 (24,7%), el 73,8% está fugado a junio de 2010. Quienes aun no estaban fugados en diciembre 2009 (75,3%), el 52,3% está fugado a junio de 2010. En el Clúster de Leales: Quienes efectivamente estaban fugados en diciembre 2009 (4,2%), el 60,7% está fugado a junio de 2010. Quienes aun no estaban fugados en diciembre 2009 (95,8%), el 24,43% está fugado a junio de 2010.

En estos últimos 2 clústeres se observa que los leales se mantienen transando con muy buenos niveles (comparativamente). El caso crítico es el segmento de alto potencial, el que es muy inestable en cuanto a transacciones.

En las siguientes tablas se muestran los resultados de la evaluación de los demás tamaños de empresa. En negrita se muestran los principales resultados en cuando a la transaccionalidad. Cabe destacar que los resultados de estos segmentos son superiores a los mencionados en el sector microempresarial (en cuanto a certeza de evaluación de transaccionalidad de proveedores).

Tabla 41: Evaluación al año 2010 (Base 2009) (Empresas Pequeña)

	Candidatos de Fuga		Nuevos		Esporádicos		Alto Potencial		Leales	
TOTAL	4.351		3.284		2.378		3.453		3.082	
<i>Estado Transaccional Diciembre 2009</i>										
Fugado	1.285	29,5%	1.356	41,3%	379	15,9%	103	3,0%	45	1,5%
Transando	3.066	70,5%	1.928	58,7%	1.999	84,1%	3.350	97,0%	3.037	98,5%
<i>Estado Transaccional Proveedores Transando 2009 a Junio 2010</i>										
Fugado	1.418	46,2%	1.228	63,7%	724	36,2%	483	14,4%	170	5,6%
Transando	1.648	53,8%	700	36,3%	1.275	63,8%	2.867	85,6%	2.867	94,4%
<i>Estado Transaccional Proveedores Fugados 2009 a Junio 2010</i>										
Fugado	918	71,4%	1.135	83,7%	237	62,5%	68	66,0%	30	66,7%
Transando	367	28,6%	221	16,3%	142	37,5%	35	34,0%	15	33,3%
<i>Estado Transaccional TODOS a Junio 2010</i>										

Fugado	2.336	53,7%	2.363	72,0%	961	40,4%	551	16,0%	200	6,5%
Transando	2.015	46,3%	921	28,0%	1.417	59,6%	2.902	84,0%	2.882	93,5%

(Fuente: Elaboración Propia)

El 71,4% de los proveedores candidatos a fuga, fugados a diciembre 2009 permanece fugado a junio de 2010. El 83,7% de los proveedores nuevos, fugados en diciembre 2009 está fugado a junio de 2010

El 85,6% de los proveedores de alto potencial, transando a diciembre 2009 permanece transando a junio de 2010 El 94,4% de los proveedores leales, transando a diciembre 2009 permanece transando a junio de 2010

Tabla 42: Evaluación al año 2010 (Base 2009) (Empresas Mediana)

	Fugados		Candidatos a Fuga		Esporádicos		Alto Potencial		Leales	
TOTAL	702		942		1.222		599		879	
<i>Estado Transaccional Diciembre 2009</i>										
Fugado	418	59,5%	95	10,1%	144	11,8%	30	5,0%	2	0,2%
Transando	284	40,5%	847	89,9%	1.078	88,2%	569	95,0%	877	99,8%
<i>Estado Transaccional Proveedores Transando 2009 a Junio 2010</i>										
Fugado	129	45,4%	386	45,6%	289	26,8%	115	20,2%	22	2,5%
Transando	155	54,6%	461	54,4%	789	73,2%	454	79,8%	855	97,5%
<i>Estado Transaccional Proveedores Fugados 2009 a Junio 2010</i>										
Fugado	319	76,3%	69	72,6%	100	69,4%	24	80,0%	2	100,0%
Transando	99	23,7%	26	27,4%	44	30,6%	6	20,0%	-	0,0%
<i>Estado Transaccional TODOS a Junio 2010</i>										
Fugado	448	63,8%	455	48,3%	389	31,8%	139	23,2%	24	2,7%
Transando	254	36,2%	487	51,7%	833	68,2%	460	76,8%	855	97,3%

(Fuente: Elaboración Propia)

El 76,3% de los proveedores fugados a diciembre 2009 permanece fugado a junio de 2010. El 72,6% de los proveedores candidatos a fuga, fugados en diciembre 2009 está fugado a junio de 2010

El 79,8% de los proveedores de alto potencial, transando a diciembre 2009 permanece transando a junio de 2010 El 97,5% de los proveedores leales, transando a diciembre 2009 permanece transando a junio de 2010

Tabla 43: Evaluación al año 2010 (Base 2009) (Empresas Grandes)

	Candidatos a Fuga		Esporádicos		Alto Potencial		Leales	
TOTAL	339		664		559		852	
Estado Transaccional Diciembre 2009								
Fugado	105	31,0%	166	25,0%	26	4,7%	2	0,2%
Transando	234	69,0%	498	75,0%	533	95,3%	850	99,8%
Estado Transaccional Proveedores Transando 2009 a Junio 2010								
Fugado	116	49,6%	168	33,7%	61	11,4%	21	2,5%
Transando	118	50,4%	330	66,3%	472	88,6%	829	97,5%
Estado Transaccional Proveedores Fugados 2009 a Junio 2010								
Fugado	78	74,3%	123	74,1%	12	46,2%	1	50,0%
Transando	27	25,7%	43	25,9%	14	53,8%	1	50,0%
Estado Transaccional TODOS a Junio 2010								
Fugado	194	57,2%	291	43,8%	73	13,1%	22	2,6%
Transando	145	42,8%	373	56,2%	486	86,9%	830	97,4%

(Fuente: Elaboración Propia)

El 74,3% de los proveedores candidatos a fuga, fugados a diciembre 2009 permanece fugado a junio de 2010.

El 88,6% de los proveedores de alto potencial, transando a diciembre 2009 permanece transando a junio de 2010. El 97,5% de los proveedores leales, transando a diciembre 2009 permanece transando a junio de 2010.

A la luz de estos resultados se observa que el modelo entrega resultados satisfactorios, que la agrupación de proveedores es buena y que permite predecir con anterioridad los proveedores con mayor tendencia a la fuga.

6.1. Aplicación de modelo de clasificación

La aplicación del modelo de clasificación tiene que ver con determinar en qué categoría debe estar un proveedor no clasificado. A continuación se presenta la aplicación del modelo en una base de proveedores activos en transacción durante 2010, pero que no hayan transado durante 2009, es decir proveedores nuevos. Se empleará la base original (base de proveedores transando durante 2009) como base de entrenamiento.

Tabla 44: Evaluación Proveedores Nuevos 2010

CLUSTER	Micro	Pequeña	Mediana	Grande
Fugados	367		140	
Candidatos a Fuga	5.319	1.330	504	101
Nuevos	5.641	850		
Esporádicos		607	129	141
Alto Potencial	2.441	393	23	103
Leales	2.082	150	15	16
TOTAL	15.850	3.330	811	361

(Fuente: Elaboración Propia)

La tabla 43 entrega el resultado de la clasificación de las 20.352 empresas que han transado en 2010 (y que no se encuentran presentes en 2009). Corresponden a proveedores nuevos que no han sido clasificados.

Nuevamente es importante considerar que el hecho de que no haya proveedores por ejemplo en el cluster “esporádicos” de la microempresa, no implica que no existan, sino que en el modelo, no alcanzan a representar un conglomerado cuyas características significativamente representen la descripción aquí abordada.

6.2. Aplicaciones del modelo:

Utilizando la información proporcionada en este modelo es posible realizar diversas acciones focalizadas por segmento. Es importante considerar que por antecedentes manifestados a lo largo del estudio es posible inferir que este trabajo apunta directamente a los dos principales objetivos estratégicos actuales de la Dirección ChileCompra (Satisfacción de usuarios y *match* oferta-demanda).

Una de las principales preocupaciones del Área de Desarrollo Empresarial es la focalización de sus acciones, que, dado los escasos recursos, requieren alta eficiencia en términos de impacto. Es en este sentido en que se recomienda focalizar las acciones en el segmento microempresarial, primero porque en términos de volumen, representan sobre el 72% de los proveedores, y demás porque son los más susceptibles a la fuga transaccional del sistema.

En la tabla siguiente se observa la cantidad de proveedores transando a junio de 2010, incluidos en el modelo. Se observa que el sector microempresarial posee un 56% de proveedores que cumplen criterio se fuga, lo que valida la afirmación del párrafo anterior.

Tabla 45: Proveedores base 2009 transando a junio 2010

CLUSTER	Micro	Pequeña	Mediana	Grande	TOTAL
Fugados	1.547		254		1.801
Candidatos a Fuga	4.575	2.015	487	145	7.222
Nuevos	2.440	921			3.361
Esporádicos		1.417	833	373	2.623
Alto Potencial	5.717	2.902	460	486	9.565
Leales	11.903	2.882	855	830	16.470
TOTAL	26.182	10.137	2.889	1.834	41.042
Porcentaje Fuga	56%	39%	33%	24%	50%

(Fuente: Elaboración Propia)

Los proveedores de los otros segmentos (pequeña, mediano, grande) también pueden ser sujetos de actividades, pero en régimen masivo. Eventualmente a través de invitaciones a seminarios y correos electrónicos regulares con información relevante del sistema de Mercado Público.

Lo que sigue es una serie de propuestas para implementar durante el segundo semestre de 2010 utilizando de manera directa o indirecta los resultados de este estudio:

Capacitaciones:

En el ámbito de las capacitaciones a proveedores, las cuales principalmente se ejecutan en la Red de Centros de Emprendimiento, se sugiere realizar una actualización de los contenidos focalizándolos en el proveedor objetivo.

Por ejemplo, para los proveedores ubicados en los segmentos “fugados” y “candidatos a fuga”, enfocar las capacitaciones en lograr que el proveedor se reencante con el sistema, a través del análisis de casos de éxito de proveedores (definidos eventualmente a través de movilidad entre los mismos clusters definidos acá).

En lo que respecta a los proveedores nuevos, actualizar el modelo vigente de capacitación en Búsqueda de Oportunidades de Negocios, el cual es impartido semanalmente en la red.

Y por último, los proveedores en los segmentos “esporádicos” y de “alto potencial”, aprovechar su nivel de aprendizaje en el sistema y buscar su fidelización a través de la capacitación en programas más avanzados, especialmente enfocados en la nueva tendencia de Mercado Público, orientada a la búsqueda de oportunidades de

negocios on line utilizando herramientas de inteligencia de negocios (<http://analiza.mercadopublico.cl>).

Match Oferta-Demanda:

La principal acción de la Dirección ChileCompra en este ámbito es la prevención de licitaciones “no exitosas” es decir adjudicadas con menos de 3 ofertas. Es en este sentido en donde se puede encontrar una de las mejores aplicaciones del modelo de comportamiento de proveedores.

Se propone realizar un catastro de rubros por región, buscando cuales son los sectores industriales en donde el mercado público no logra cumplir con sus objetivos de eficiencia.

Con esto se puede contactar, ya sea a través de la red de Centros de Emprendimiento o campañas de mailing masivo (controladas por el sistema CRM de la Dirección ChileCompra), a los proveedores esporádicos o con alto potencial de oferta para que aprovechen esas oportunidades de negocios y aumenten su participación en este mercado.

Esta es una medida de alto impacto en lo referido a potenciar la eficiencia del sistema al lograr contar con procesos de licitación adjudicados bajo los óptimos de eficiencia según competencia.

En el caso de los proveedores del sector microempresarial, se sugiere la difusión de oportunidades de negocio del tipo L1 y LE²². Para proveedores de otros segmentos, incluir las licitaciones tipo LP.

Fidelización:

Una de las características de este mercado es el hecho de que no existen barreras de entrada ni de salida. Es por ello que la fuga transaccional de proveedores puede darse por razones relacionadas con mercado (eficiencia), compradores (transparencia, probidad) o simplemente por incentivos de mercados externos.

Existen segmentos de proveedores esporádicos y con alto potencial de oferta. Estos se caracterizan (especialmente en el sector micro), por no estar registrados ni vigentes en el registro de proveedores del Estado (ChileProveedores).

²² Una licitación del tipo L1 es una licitación cuyo monto oscila entre 0 y 100 UTM. Una licitación del tipo LE es una licitación cuyo monto oscila entre las 100 UTM y las 1.000 UTM. Una licitación del tipo LP, es una licitación cuyo monto supera las 1.000 UTM (aprox 37.000.000 de pesos).

Actualmente cerca de 40.000 empresas se mantienen vigentes en el registro, con posibilidad de acceder a múltiples beneficios entre los que destacan un directorio de empresas, y un catálogo de productos con orientación a la transacción internacional.

Entonces se propone una campaña intensa de fidelización de estos proveedores potenciales oferentes en el registro pues al insertarlos en él, no solo recibirán beneficios sino que además al tener que cancelar una cuota²³, se estima que existe una especie de barrera de salida pues el registro puede considerarse como una inversión.

Atención Personalizada en la Red de Centros de Emprendimiento:

Los Centros de Emprendimiento ChileCompra desde 2008 han sido el puntal del plan de acceso de la institución, facilitando el acceso al Mercado de las Compras Públicas a más de 20.000 proveedores en su mayoría del sector microempresarial. Estos Centros de Emprendimiento, comenzaron con diversos programas buscando el desarrollo y fortalecimiento de empresas a través de alianzas con instituciones como Sercotec, Fosis, Corfo, etc.

Actualmente la red cuenta con cobertura nacional, con más de 30 Centros de Emprendimiento. Cuenta con profesionales con experiencia capaces de orientar en procesos licitatorios a los proveedores que lo necesiten.

Se propone entonces, utilizar este “brazo” de la Dirección como un punto de contacto entre los proveedores locales con mayor probabilidad de fuga, a través del contacto personalizado con el Director Ejecutivo del Centro de Emprendimiento.

Actualmente los Centros de Emprendimiento cuentan con un sistema de control de gestión on line, el cual eventualmente podría incluir dentro de sus indicadores los datos de categorización de proveedores descritos en este informe.

Comentarios generales sobre estas propuestas:

Es importante destacar que para el segundo semestre de 2010 serán implementadas estas acciones dentro del plan de actividades del Área de Desarrollo Empresarial, como complemento al plan de trabajo anual vigente.

Se ha determinado focalizar los esfuerzos en el sector microempresarial considerando los segmentos candidatos a fuga, nuevos y de alto potencial de oferta. Excluyendo de los esfuerzos a los proveedores fugados y los en categoría de leales.

En el caso de los fugados pues actualmente las estrategias estaban enfocadas intentar contactar a proveedores inactivos desde más de un año hacia atrás obteniendo

²³ <http://www.chileproveedores.cl/DNN/Tarifas.aspx>

resultados poco alentadores. En el caso de los proveedores leales, se asume que son proveedores que regularmente transan en Mercado Público, por ende solo debiesen entrar en el target estratégico una vez que su comportamiento transaccional disminuya en volumen.

Como último alcance a las aplicaciones del modelo, a modo de recomendación se sugiere la creación de una figura responsable de hacerse cargo de la Fidelización de Clientes Proveedores; eventualmente a través de un cargo dentro de la estructura del Área de Desarrollo Empresarial o como productos dentro de la misma. En rigor se busca con esta idea mantener en la mirada estratégica de la institución el cuidado de los clientes.

7. CAPITULO 7: CONCLUSIONES

7.1. Conclusiones

Sobre el modelo:

El modelo aquí descrito le entrega al Área de Desarrollo Empresarial de la Dirección ChileCompra una herramienta para la toma de decisiones a través de la caracterización de sus clientes en base a indicadores y criterios de negocio. Esta mejora en las acciones del área puede traer múltiples beneficios a la institución, dentro de los cuales se puede destacar el aumento de los proveedores transando en el sistema. Al identificar a los proveedores con probabilidad de deserción es posible retenerlos y cumplir con uno de los grandes objetivos del sistema de Mercado Público. Garantizar la eficiencia en el mercado de las compras públicas.

Otro ámbito es la eficiencia, la cual no solo se manifiesta en la adjudicación de procesos de licitación, sino que además el Área de Desarrollo Empresarial al focalizar sus recursos identificando segmentos con tendencia a la deserción, genera políticas más efectivas. Otro beneficio que la Dirección recibe es la mejora continua de los servicios y productos diseñados para la cartera de proveedores, logrando así cumplir de mejor manera con los estándares de calidad y equidad en el acceso al sistema. Esto puede entregar una mejora en la percepción de la calidad del servicio, lo que impacta directo en la Satisfacción de los usuarios, otra meta institucional de la Dirección ChileCompra.

En general se ha logrado construir un modelo de comportamiento de proveedores en base a estados en el sistema. Se le ha entregado al Área de Desarrollo Empresarial un set de criterios y distinciones para aumentar el nivel de conocimiento sobre sus clientes.

Se ha logrado identificar quienes son los proveedores más propensos a desertar del sistema, lo que representa un primer paso a la prevención del fenómeno. En rigor se ha logrado determinar que ciertas variables transaccionales pueden originar predictores de comportamiento, así las estrategias actuales pueden ser redefinidas. Entre un 80 y un 90% de los proveedores que se espera fuguen del sistema efectivamente lo hacen. Rangos de efectividad similares se obtienen con los proveedores que se estima son leales.

Actualmente uno de los focos está en la inscripción de proveedores en el sistema, lo que ha originado que durante los últimos años el recambio de proveedores transando sea muy elevado. Por ende, ahora se propone una mirada desde el punto de vista del cuidado de los proveedores transando en la actualidad.

En general se ha logrado validar la hipótesis planteada en los alcances de este estudio al manifestar que la información transaccional es un buen predictor de comportamiento de proveedores en el sistema de Mercado Público.

Sobre las técnicas empleadas

El trabajo realizado refleja las bondades de la aplicación de técnicas de inteligencia de negocios en el sistema. Se ha logrado extraer valor de la información transaccional de la plataforma a través del uso de la minería de datos.

Se demuestra que el proceso de análisis que intenta buscar relaciones desconocidas entre datos, llamado minería de datos es un proceso práctico, que ayuda a explicar fenómenos y realizar predicciones con ellos de manera rápida y eficiente, sin necesidad de ser un experto en el tema a investigar. En este caso se trata de un estudio que entrega interesantes y muy valorados resultados en cuanto al análisis del comportamiento de proveedores sin contar con antecedentes previos en lo que se refiere a la generación de conocimiento.

Se utilizaron algoritmos clásicos en la literatura, empleados en estudios similares en otros campos, entregando información consistente con la observación práctica y experiencia del equipo del Área de Desarrollo Empresarial.

Comentarios finales:

Como conclusión final, se puede afirmar que se cumplió con los productos acordados con la Dirección ChileCompra, dentro de los tiempos comprometidos, y lo que es más importante aún con los objetivos diseñados.

Se desarrolló un modelo de comportamiento de proveedores en el sistema de Mercado Público como apoyo para la toma de decisiones del Área de Desarrollo Empresarial de la Dirección ChileCompra. Modelo que ha sido valorado y bien recibido por el equipo de profesionales de la institución, quienes han depositado su confianza en él. Esto se manifiesta a través de la aplicación de estos criterios en la selección de proveedores objetivo de actividades durante lo que resta de 2010.

Por otro lado, el desarrollo de este modelo abre las puertas a mejoras en el mismo modelo y a nuevos desarrollos en cuanto a la extracción de valor desde el sistema, lo que representa un interesante desafío para el Área, al entregarle nuevas perspectivas enmarcadas en la línea de la visión de la institución al ofrecer servicios de calidad y originando avances en la modernización de la gestión y en la implementación de políticas públicas.

7.2. Mejoras al modelo, trabajo futuro y comentarios generales del sistema

En esta sección se realizarán algunos comentarios a modo de sugerencia de mejora al modelo y al sistema de Mercado Público en general, que se estima aportarán mayor valor al desarrollo de conocimiento en el sistema de Mercado Público.

Sobre las variables empleadas

En general las variables transaccionales posibles de implementar han sido recogidas en el modelo, pero considerar solo este tipo de variables sesga los resultados. Existen variables que pudiesen ser incluidas. Tales como:

- **Participación en ventas del Mercado Público:** Corresponde a la evaluación del porcentaje de las utilidades de un proveedor provenientes del Mercado Público. Se estima que esta variable pudiese aportar valor predictivo a manifestar la necesidad de la empresa de mantenerse transando en el mercado público.
- **Nivel de capacitación:** Corresponde a internalizar en la medición el nivel de capacitación de los usuarios proveedores en cuanto al uso de la plataforma. Especialmente enfocado en los clústeres de proveedores nuevos que presentan relativamente altos valores de ofertas rechazadas con respecto a los demás. Se estima que esto se origina por el desconocimiento de cómo realizar buenas ofertas a procesos de licitación publicados en el sistema.

Sobre la actualización de los modelos:

Se propone el desarrollo de metodologías para la actualización del modelo diseñado. Estas metodologías debiesen contemplar la actualización de las variables que determinan el modelo y los parámetros del mismo. De esta forma con las actualizaciones es posible incorporar la dimensión dinámica del mercado capturando información del patrón de fuga transaccional en el tiempo.

Eventualmente con información en el tiempo sobre la composición de segmentos y en general el movimiento de los proveedores se puede lograr generar modelos de clustering temporal, data mining dinámico, ciclo de vida o la detección de cambios en el comportamiento de clientes (los cuales incluso pueden contrastarse con la demanda, es decir la dinámica presupuestaria del estado).

Una propuesta en este sentido es la construcción de un data mart de proveedores, que contenga como principales medidas las ofertas y el comportamiento transaccional de los proveedores en el tiempo.

Sobre las variables objetivo de la clasificación:

Durante el desarrollo de la investigación, se aplicó un modelo de clasificación con árboles de decisión, el cual consideró como variable objetivo la clasificación realizada previamente. Se propone el análisis de modelos de este tipo pero enfocado en otras variables, tales como la tasa de éxito y el tiempo de permanencia en el sistema.

Estos modelos originarían predicciones por excelencia, en el caso de la tasa de éxito, eventualmente podría determinar la probabilidad de éxito de un proveedor en un proceso de licitación sin considerar aún los criterios netos de éxito originados por las bases de licitación, lo que originaría importantes beneficios a los proveedores al disponer esta información con la finalidad de que puedan optimizar sus recursos y aumentar sus probabilidades de éxito en el mercado.

Sobre la cantidad de proveedores en el sistema.

Se sugiere realizar estudios analizando el número de proveedores que se requiere para sectores específicos. Por ejemplo, analizar sectores industriales por región del país, evaluando si los proveedores actuales (oferta) logran satisfacer la demanda actual. Esto impacta directamente en las acciones definidas en el ámbito del match oferta demanda del Mercado Público.

Sobre las causas de la fuga

En el capítulo 2 de este estudio se enunciaron algunas de las probables causas de la fuga del sistema por parte de los proveedores. No fue objetivo el análisis particular de estas causas, pero es conveniente sugerir metodologías de evaluación de las mismas o propuestas de evaluación de las mismas con la finalidad de reducirlas.

Existen causas de fuga en el ámbito de las licitaciones, las cuales principalmente se dan en la construcción de las licitaciones, por ejemplo la publicación de Oportunidades de negocios en plazos no adecuados para el normal proceso de oferta, los criterios ambiguos, la exigencia de marcas específicas, etc. En general esto se contrarresta con la evaluación de los compradores y el monitoreo de estas variables. Eventualmente se pueden diseñar criterios de riesgo de licitaciones.

Las causas enfocadas en el financiamiento de proveedores, como por ejemplo la exigencia de boletas de garantía y el sistema de pagos de los organismos públicos (hoy en día el 50% de los reclamos de proveedores se enfoca en problema de pago de servicios públicos, especialmente del sector municipal²⁴), se están combatiendo a

²⁴ Información extraída directamente de las bases de datos operacionales.

través de exigencias a los mismos organismos y la oferta de alternativas de financiamiento, tales como confirming, ordering, los certificados de fianza, etc.

Sobre el criterio de licitaciones exitosas.

El criterio actual de licitaciones exitosas consta de la evaluación de la oferta, es decir, si una licitación cuenta con 3 o más ofertas en su proceso (Milgrom, 2004), es considerada exitosa. Este criterio solo obedece a mantener el nivel de oferta en el sistema y garantizar eficiencia en la adjudicación.

Se sugiere incorporar criterios de construcción de licitaciones, plazos (por ejemplo adjudicación en momentos definidos en bases), etc. La finalidad es que se pueda evaluar una licitación exitosa en todos sus aspectos, no solamente en lo que se refiere a la “adjudicación exitosa”.

8. CAPITULO 8: BIBLIOGRAFÍA

8.1. Bibliografía

Acosta, Rolando, y otros. 2008. *Empleo de Minería de Datos en la predicción de Diabetes*. La Habana : VII Congreso Internacional de Informática en la Salud, 2008.

Aluja, Tomas. 2001. La Minería de datos, entre la estadística y la inteligencia artificial. La Rioja : Universidad de la Rioja, 2001, Vol. 25, págs. 479-498.

Briones, Diego. 2002. *Modelo de Predicción de Fuga de Clientes para Banco Estado*. Santiago : s.n., 2002.

Britos, P., y otros. 2008. *Explotación de Información Aplicada a Inteligencia Criminal en Argentina*. Buenos Aires : Centro de Ingeniería del Software e Ingeniería del Conocimiento, 2008.

ChileProveedores. 2010. Registro Nacional de Proveedores. *ChileProveedores*. [En línea] 2010. <http://www.chileproveedores.cl>.

CRISP-DM. 2010. Cross Industry Standard Process for Data Mining. *CRISP-DM*. [En línea] 2010. <http://www.crisp-dm.org/>.

Dirección ChileCompra. 2008. *Plan Estratégico 2008 - 2010*. Santiago : s.n., 2008. pág. 80.

Dirección ChileCompra. 2009. *Balance Gestión 2009 y Plan Estratégico 2010 - 2012*. Dirección ChileCompra. Santiago : s.n., 2009. pág. 80, Balance Institucional.

Dirección ChileCompra. 2010a. Centros de Emprendimiento. *Centros de Emprendimiento*. [En línea] 2010. http://www.chilecompra.cl/centros_de_emprendimiento.html.

Dirección ChileCompra. 2010b. Consultas y Ayudas. *Dirección ChileCompra*. [En línea] 2010. http://www.chilecompra.cl/secciones/consultas_y_ayudas/Compradores/06_Orden_Compra_compradores.html.

Dirección ChileCompra. 2010c. Dirección de Compras y Contratación Pública. *ChileCompra*. [En línea] 2010. <http://www.chilecompra.cl>.

Dirección ChileCompra. 2010d. Mercado Público. *Mercado Público*. [En línea] 2010. <http://www.mercadopublico.cl>.

From data mining to knowledge discovery in databases. **Fayyad, Usama, Piatetsky-shapiro, Gregory y Smyth, Padhraic. 1996.** s.l. : AI Magazine, 1996, American Association for Artificial Intelligence. ISBN: 0738-4602-1996.

- Gutierrez, Patricio. 2008.** *Caracterización de la Población Carcelaria en Argentina Mediante la Aplicación de Minería de Datos para la Prevención de Hechos Delictivos.* s.l. : Instituto Tecnológico de Buenos Aires, 2008.
- Hand, David J., Mannila, Heikki y Smyth, Padhraic. 2001.** *Principles of Data Mining.* s.l. : The MIT, 2001. ISBN: 0-262-08290-X.
- Huete, L. M. 1996.** *Revitalizando los Servicios: Tres enfoques de gestión que serán comunes en las empresas del siglo XXI.* Barcelona : Folio, 1996. ISBN: 84-413-0606-0.
- Hurtado, Alberto. 2004.** *Moral Social: Obra Póstuma de San Alberto Hurtado, S.J. .* Segunda Edición . Santiago : Universidad Católica de Chile, 2004. ISBN: 956-14-0752-3.
- KDnuggets. 2010.** Data Mining Commutiny's top resource since 1997 for Data Mining and Analytics Software. *KDnuggets.* [En línea] 2010. <http://www.kdnuggets.com/>.
- Macias, Miguel. 2008.** *Técnicas de Minería de Datos para la Retención de Clientes en el Sector Asegurador.* s.l. : Comisión Nacional de Seguros y Finanzas, 2008.
- Milgrom, Paul. 2004.** *Putting Auction Theory to Work.* Cambridge : Cambridge University Press, 2004. ISBN: 0-521-55184-6.
- Ministerio de Hacienda. 2003.** Ley de bases sobre contratos administrativos de suministro y prestación de servicios. *Ley Número 19.886.* [Publicación Digital]. s.l., Santiago, Chile : Congreso Nacional, 2003.
- Miranda, Jaime, Rey, Pablo y Weber, Richard. 2005.** *Predicción de Fugas de Clientes para una institución financiera mediante Support Vector Machines.* s.l. : Revista Ingeniería de Sistemas, 2005. Vol. XIX.
- Molina, José y Garcia, Jesus. 2004.** *Técnicas de Analisis de Datos.* Madrid : Universidad de Madrid, 2004.
- Observatorio Empresas. 2010.** Observatorio Empresas. *Observatorio Empresas.* [En línea] 2010. <http://www.observatorioempresas.cl>.
- Quinlan, J. Ross. 1979.** *Discovering rules from large collections of examples.* Edimburgo : Edimburgo University Press, 1979.
- Quinlan, J. Ross. 1993.** *C4.5: Programs for machine learning.* Sydney : Morgan Kaufmann, 1993. ISBN:1-55860-238-0.

Segovia, Carolina, Aburto, Luis y Goic, Marcel. 2005. Caracterización del Proceso de Fuga de Clientes Utilizando Información Transaccional. *Penta Analytcs*. [En línea] 2005. VI Congreso Chileno de Investigación Operativa. <http://www.penta.cl/analytics/pages/publicaciones.htm>.

SEMMA. 2010. SAS Enterprise Miner SEMMA. [En línea] 2010. <http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html>.

Tang, ZhaoHui y MacLennan, Jamie. 2005. *Data Mining with SQL Server 2005*. s.l. : Willey, 2005. ISBN: 0471462616.

Waikato. 2010. Weka 3: Data Mining Software in Java. *University of Waikato*. [En línea] 2010. <http://www.cs.waikato.ac.nz/ml/weka/>.

Weber, Richard. 2000. *Data Mining en la Empresa Utilizando Tecnologías Inteligentes*. s.l. : Revista Ingeniería de Sistemas, 2000.

Weiss, Sholom M. y Indurkha, Nitin. 1998. *Predictive Data Mining*. San Francisco : Morgan Kauffman, 1998. ISBN: 1-55860-403-0.

Witten, Ian y Frank, Eibe. 2005. *Data Mining: Practical Machine Learning tools and techniques*. San Francisco : Morgan Kaufmann, 2005. ISBN 0-12-088407-0.

9. CAPITULO 9: ANEXOS

9.1. Anexo A: Tipos de Contratación

Existen cuatro tipos de contratación en el sistema de Mercado Público. Los cuales se explican brevemente a continuación:

Convenio Marco: El Convenio Marco es un acuerdo celebrado previa licitación pública, por la Dirección de Compras y Contratación Pública con un proveedor determinado. En el se establecen precios, condiciones de compra u otras especificaciones del producto o servicio. Es un Catálogo Electrónico que está disponible en el portal ChileCompra para los usuarios del sistema. Es importante destacar que existe obligatoriedad de uso para Organismos Públicos (uso obligatorio para Fuerzas Armadas y de Orden y Seguridad Pública, salvo excepciones definidas en el Reglamento Especial) y uso voluntario para Municipios

Licitación Pública: Procedimiento administrativo de carácter concursal mediante el cual la administración realiza un llamado público convocando a los interesados para que, sujetándose a las bases fijadas, formulen propuestas (ofertas), entre las cuales seleccionará y aceptará la más conveniente. Cualquier persona puede presentar ofertas y el llamado es a través de sistema de Información o bien la entidad licitante podrá publicarlo por medio de uno más avisos.

Licitación Privada: Procedimiento administrativo de carácter concursal, previa resolución fundada que lo disponga, mediante el cual la Administración invita a determinadas personas para que, sujetándose a las bases fijadas, formulen propuestas, de entre las cuales seleccionará y aceptará las más conveniente.

Trato o contratación directa: Procedimiento de contratación que, por la naturaleza de la negociación que conlleva, deba efectuarse sin la concurrencia de los requisitos señalados para la licitación o propuesta pública y para la privada

El uso de las licitaciones privadas y tratos directos están normados en la Ley de Compras a través del artículo 8, el cual se enuncia a continuación.

- Procederá la licitación privada o el trato o contratación directa en los casos fundados que a continuación se señalan:
 - Si en las licitaciones públicas respectivas no se hubieren presentado interesados.
 - Si se tratara de contratos que correspondieran a la realización o terminación de un contrato que haya debido resolverse o terminarse anticipadamente por falta de cumplimiento del contratante u otras causales y cuyo remanente no supere las 1.000 unidades tributarias mensuales
 - En casos de emergencia, urgencia o imprevisto.
 - Si sólo existe un proveedor del bien o servicio.

- Si se tratara de convenios de prestación de servicios a celebrar con personas jurídicas extranjeras que deban ejecutarse fuera del territorio nacional
- Si se trata de servicios de naturaleza confidencial
- Naturaleza de la negociación
- Contratación igual o inferior a UTM 100.

Criterios de Naturaleza de la negociación:

- Si se requiere contratar la prórroga de contrato o contratar servicios conexos
- Gastos de representación
- Cuando pueda afectarse la seguridad e integridad personal de autoridades, siendo necesario contratar con un proveedor probado que asegure discreción y confianza
- Consultorías cuyas materias se encomienden en consideración a facultades de Proveedor que dará el servicio (artículo 105 y siguientes del Reglamento).
- Contratación sólo pueda realizarse con proveedores titulares de derechos propiedad intelectual, industrial, etc.
- Cuando por la magnitud e importancia de la contratación es indispensable recurrir a un proveedor determinado en la razón de la confianza y seguridad que se derivan de su experiencia comprobada, y siempre que se estime fundadamente que no existen otros proveedores que den dicha confianza.
- Reposición o complementación de equipamiento o servicios accesorios que deban ser compatibles con modelos, sistemas, infraestructura previamente adquirida por la entidad.
- Cuando el conocimiento público de licitación pudiese poner en riesgo el objeto y eficacia de contratación.

9.2. Anexo B: ChileProveedores

Glosario con principales definiciones del sistema (ChileProveedores, 2010).

- HÁBIL: Proveedor HABILITADO ya que NO presenta alguna de las causales descritas en el Artículo 92 del Reglamento de la Ley de Compras Públicas N°19.886.
- INHÁBIL: Proveedor NO HABILITADO ya que presenta alguna(s) causal(es) descritas en el Artículo 92 del Reglamento de la Ley de Compras Públicas N°19.886.

Artículo 16 Ley de Compras Públicas: "Se inscribirán en el Registro, todas las personas naturales y jurídicas, chilenas y extranjeras que no tengan causal de inhabilidad para contratar con los organismos del Estado."

Artículo 92 del Reglamento de la ley de Compras Públicas N°19.886. "Estarán inhabilitados para inscribirse en el Registro de Proveedores, los que se encuentren en alguna de las siguientes circunstancias:"

Haber sido condenado por cualquiera de los delitos de cohecho contemplados en el título V del Libro Segundo del Código Penal.

Registrar una o más deudas tributarias por un monto total superior a 500 UTM por más de un año, o superior a 200 UTM e inferior a 500 UTM por un período superior a 2 años, sin que exista un convenio de pago vigente. En caso de encontrarse pendiente juicio sobre la efectividad de la deuda, esta inhabilidad regirá una vez que se encuentre firme o ejecutoriada la respectiva resolución.

Registrar deudas previsionales o de salud por más de 12 meses por sus trabajadores dependientes, lo que se acreditará mediante certificado de la autoridad competente.

La presentación al Registro Nacional de Proveedores de uno o más documentos falsos, declarado así por sentencia judicial ejecutoriada.

Haber sido declarado en quiebra por resolución judicial ejecutoriada.

Haber sido eliminado o encontrarse suspendido del Registro Nacional de Proveedores por resolución fundada de la Dirección de Compras.

Haber sido condenado por prácticas antisindicales o infracción a los derechos fundamentales del trabajador.

9.3. Anexo C: Área de Desarrollo Empresarial

Centros de Emprendimiento

En el marco del Programa de Acceso ChileCompra y con la finalidad de brindar apoyo constante a la micro y pequeña empresa regional, promoviendo su participación en las oportunidades de negocio que genera el mercado de las compras públicas, la Dirección ChileCompra cuenta con una red de Centros de Emprendimiento, ubicados en las distintas regiones del país, en alianza con otras instituciones públicas y privadas locales. Actualmente existen 30 Centros de Emprendimiento que al 2009 trabajan directamente con más de 16.000 proveedores (Dirección ChileCompra, 2010).

ChileProveedores

Es el registro electrónico de proveedores del Estado, el cual, reúne toda la información relevante de aquellas empresas que participan en las licitaciones estatales. El registro incluye información de los productos, servicios, proyectos y certificaciones de las empresas. Adicionalmente, contempla la información comercial, legal y financiera, que permite acreditar que la empresa cumple con los requisitos de contratación con Entidades Públicas (ChileProveedores, 2010).

Confirming

El Confirming o liquidación de facturas es una operación que deriva del factoring, que consiste en el pago anticipado a Proveedores de las facturas cedidas a un Inversionista, previa información y confirmación de pago de la Entidad pagadora al mismo inversionista. En esta figura, el Inversionista paga el monto de factura con tasa de descuento ofertada al Proveedor, y el organismo de la Administración del Estado, en adelante, la Entidad pagadora o Entidad, paga al inversionista el total de la factura (Dirección ChileCompra, 2009).

Ordering

Producto ofrecido por instituciones financieras a los proveedores de ChileCompra, que permite obtener créditos para financiar un negocio respaldado por la orden de compra emitida a través del sistema (Dirección ChileCompra, 2009).

9.4. Anexo D: Criterio MVI: “Máxima Ventana de Inactividad”

El criterio de la Máxima Ventana de Inactividad de un proveedor en el sistema intenta capturar los “ciclos” de actividad que registra un proveedor, asumiendo que la inactividad es temporal y que en algún instante en el futuro el proveedor volverá a transar.

En general se trata de capturar el mayor ciclo de inactividad previo al último periodo inactivo corriente. Si el último periodo inactivo es mayor o igual al MVI, entonces se catalogará al proveedor como fugado, pues esto indica que superó su mayor ciclo de inactividad significando que probablemente no volverá.

Ilustración 35: Criterio de Fuga “Máxima Ventana de Inactividad” (MVI)

Prov	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	MAX MVI	VI Actua l	Estado
A	0	0	1	0	0	1	0	2	1	Activo
B	0	0	0	0	1	0	1	4	0	Activo
C	0	1	0	0	0	0	0	1	5	Fugado

(Fuente: Elaboración Propia)

En la ilustración anterior se observan tres ejemplos. El proveedor A posee una MVI de 2, y su último periodo inactivo es de 1, entonces este proveedor será considerado activo. El proveedor C, en cambio, posee una MVI de 1 y un último periodo de inactividad igual a 5, por ende, se considerará como fugado.

Este criterio es empleado en instituciones financieras para modelar el problema. Es un criterio que deja en manos del proveedor su definición de fuga, a diferencia del criterio SMA que la define como una regla de negocio.

Ambos criterios (SMA y MVI) modelan el mismo efecto, pero obtienen datos distintos. Esto se explica principalmente por la línea en donde se considera fugado un proveedor. El criterio SMA posee valores más altos de fuga en general pues no permite considerar como activos a proveedores inactivos en un horizonte superior a 6 meses, en cambio el criterio de MVI considera activos a proveedores que en el caso extremo tienen una MVI de 12 meses, y su última actividad se registró hace 11 meses.

9.5. Anexo E: Demostración Intervalo Óptimo para Criterio SMA.

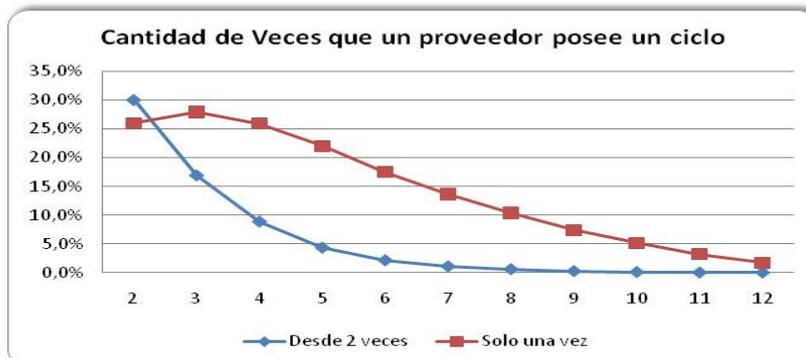
Ilustración 36: Máxima Ventana Inactiva en Periodo Activo



(Fuente: Elaboración Propia)

Una ventana inactiva en su periodo activo equivale al máximo periodo en que estuvo inactivo mientras transaba (aun considerándose fugado en algún momento). En el gráfico se observa en el eje X los periodos y en el eje Y el porcentaje de proveedores activos en 2008 que posee una ventana inactiva de ese largo. El 4,6% de los proveedores posee una ventana de inactividad en su periodo activo sobre los 12 meses. Y el 23,7% posee una ventana de inactividad sobre 7 meses. El 76,3% de los proveedores tiene una máxima ventana inactiva en su periodo menor a 6 meses.

Ilustración 37: Cantidad de veces que un proveedor posee un ciclo



(Fuente: Elaboración Propia)

En el caso de los proveedores que poseen ciclos de inactividad en su periodo activo. Se puede observar que el 30% posee más de una vez un ciclo de inactividad de 2 meses, pero solo el 2,1% de los proveedores con ciclos, posee un ciclo de 6 meses de inactividad más de una vez. La probabilidad de encontrar ciclos de 6 meses en proveedores es menor al 2,1%. Por ende, evaluar actividad en intervalo de semestre móvil es un buen indicador y candidato a predictor.

9.6. Anexo F: Implementación de Técnicas de Análisis de Datos en Weka

Anexo F.1: Algoritmo k-means

- Clase que lo implementa: *weka.clusterers.SimpleKMeans.java*
- Las opciones de configuración que posee se muestran en la siguiente tabla

Tabla 46: Opciones de Configuración k-means en Weka

Opción	Default	Descripción
numCluster	2	Número de clusters que se le pide al modelo.
Seed	10	Semilla a partir de la cual se genera el número aleatorio para inicializar los centros de los clusters

(Fuente: (Molina, y otros, 2004))

- Es importante considerar que el algoritmo admite atributos simbólicos y numéricos.
- Para obtener los centroides iniciales se emplea un número aleatorio obtenido a partir de la semilla empleada.
- No se estandarizan los argumentos, sino que se normalizan.

Anexo F.2: Algoritmo EM

- Clase que lo implementa: *weka.clusterers.EM.java*.
- Las opciones de configuración que posee se muestran en la siguiente tabla.

Tabla 47: Opciones de Configuración EM en Weka

Opción	Default	Descripción
numCluster	-1	Número de clusters que se le pide al modelo. Si el número es -1, el algoritmo determinará automáticamente el número de clusters
maxIteration	100	Número máximo de iteraciones del algoritmo si este no ha convergido antes
Debug	False	Muestra información sobre el proceso de clustering.
Seed	100	Semilla a partir de la cual se genera el número aleatorio para inicializar los centros de los clusters
minStdDev	e^{-6}	Desviación típica mínima admisible en las distribuciones de densidad.

(Fuente: (Molina, y otros, 2004))

Anexo F.3: Algoritmo J48

- Clase que lo implementa: *weka.classifiers.j48.java*
- Las opciones de configuración que posee se muestran en la siguiente tabla

Tabla 48: Opciones de Configuración algoritmo J48

Opción	Default	Descripción
minNumObj	2	Número mínimo de instancias por hoja (para realizar poda)
saveInstanceData	False	Una vez finalizada la creación del árbol de decisión se eliminan todas las instancias que se clasifican en cada nodo, que hasta el momento se mantenían almacenadas.
unpruned	False	En caso de no activar esta opción, se realiza la poda del árbol.
binarySpills	False	Con los atributos nominales también no se divide el nodo en dos ramas.
subTreeRaising	True	Permitir podado con proceso de sub arboles.
confidenceFactor	0,25	Factor de confianza para el podado del árbol.

(Fuente: (Molina, y otros, 2004))

- Es importante considerar que el algoritmo J48 es la implementación en Weka del algoritmo C4.5
- Este algoritmo permite atributos simbólicos y numéricos

Anexo F.4: Algoritmo PART

- Clase que lo implementa: *weka.classifiers.j48.PART.java*
- Las opciones de configuración que posee se muestran en la siguiente tabla

Tabla 49: Opciones de Configuración algoritmo PART

Opción	Default	Descripción
minNumObj	2	Número mínimo de instancias por hoja (para realizar poda)
binarySpills	False	Con los atributos nominales también no se divide el nodo en dos ramas.
confidenceFactor	0,25	Factor de confianza para el podado del árbol.

(Fuente: (Molina, y otros, 2004))

- Se puede observar que las opciones de configuración de este algoritmo corresponden a un subconjunto del algoritmo J48, ya que los cálculos de error son exactamente los mismos.
- Es importante considerar que los creadores de este algoritmo, son los mismos que lo implementaron en este software.

9.7. Anexo G: Selección de Atributos

Anexo G.1: Empresas Micro

Resultados de algoritmos de selección de atributos para las empresas tamaño micro. En negrita las variables finalmente seleccionadas.

Tabla 50: Selección de Atributos Empresas Micro

GainRatioAttributeEval		ChiSquaredAttributeEval		InfoGainAttributeEval	
Valor	Variable	Valor	Variable	Valor	Variable
0.5417	Cantidad Adjudicaciones	84420,3073	Monto Transado	1.029	Cantidad Adjudicaciones
0.4667	Monto Transado	83246,2738	Cantidad Adjudicaciones	1,0172	Monto Transado
0.42	Cantidad Órdenes de Compra	70468,4412	Cantidad Órdenes de Compra	0.9148	Cantidad Órdenes de Compra
0.3446	Tasa "Win rate"	63812,3013	Cantidad Ofertas	0.776	Tasa "Win rate"
0.3382	Cantidad Ofertas	61448,2424	Tasa "Win rate"	0.7481	Cantidad Ofertas
0.3367	Registrado ChileProveedores	47772,3289	Cantidad de meses logeado	0.5453	Cantidad de meses logeado
0.2839	Vigente ChileProveedores	36436,4758	Promedio Logeo Mensual	0.4052	Promedio Logeo Mensual
0.2143	Promedio Logeo Mensual	23951,0367	Registrado ChileProveedores	0.3367	Registrado ChileProveedores
0.1736	Estado Transaccional	21077,8638	Vigente ChileProveedores	0.2628	Vigente ChileProveedores
0.1554	Cantidad de meses logeado	12637,2925	Año Primera Interacción	0.1651	Año Primera Interacción
0.1273	Cantidad de Reclamos	11431,7788	Estado Transaccional	0.1489	Estado Transaccional
0.1195	Ofertas Rechazadas	10746,2091	Tasa "Of-Rate"	0.1185	Tasa "Of-Rate"
0.1194	Tasa "Of-Rate"	9544,0676	Ofertas Rechazadas	0.1153	Ofertas Rechazadas
0.0898	Año Primera Interacción	2332,7163	Cantidad de Reclamos	0.0259	Cantidad de Reclamos

(Fuente: Elaboración propia)

Anexo G.2: Empresas Pequeñas

Resultados de algoritmos de selección de atributos para las empresas tamaño pequeño. En negrita las variables finalmente seleccionadas.

Tabla 51: Selección de Atributos Empresas Pequeñas

GainRatioAttributeEval		ChiSquaredAttributeEval		InfoGainAttributeEval	
Valor	Variable	Valor	Variable	Valor	Variable
0.5307	Cantidad Adjudicaciones	30695,3641	Cantidad Adjudicaciones	1,1462	Cantidad Adjudicaciones
0.4387	Vigente ChileProveedores	25880,9485	Monto Transado	0.9798	Monto Transado
0.4213	Registrado ChileProveedores	24098,5505	Cantidad de Órdenes de Compra	0.9257	Cantidad de Órdenes de Compra
0.4058	Monto Transado	23940,4929	Cantidad Ofertas	0.9163	Cantidad Ofertas
0.399	Cantidad de Órdenes de Compra	16131,5893	Cantidad de meses logeado	0.719	Cantidad de meses logeado
0.3953	Cantidad Ofertas	14493,343	Promedio Logeo Mensual	0.6311	Tasa "Win Rate"
0.2974	Promedio Logeo Mensual	13437,3174	Tasa "Win Rate"	0.6149	Promedio Logeo Mensual
0.233	Tasa "Win Rate"	8458,67	Vigente ChileProveedores	0.4385	Vigente ChileProveedores
0.2121	Cantidad de meses logeado	7887,5251	Registrado ChileProveedores	0.3869	Registrado ChileProveedores
0.1823	Estado Transaccional	5681,5266	Ofertas Rechazadas	0.2318	Ofertas Rechazadas
0.1682	Ofertas Rechazadas	3851,867	Tasa "Of-rate"	0.1611	Tasa "Of-rate"
0.1444	Cantidad de Reclamos	3146,88	Año Primera Interacción	0.1376	Año Primera Interacción
0.1314	Tasa "Of-rate"	2565,0708	Estado Transaccional	0.1284	Estado Transaccional
0.0825	Año Primera Interacción	1167,4987	Cantidad de Reclamos	0.0456	Cantidad de Reclamos

(Fuente: Elaboración propia)

Anexo G.3: Empresas Medianas

Resultados de algoritmos de selección de atributos para las empresas tamaño mediano. En negrita las variables finalmente seleccionadas.

Tabla 52: Selección de Atributos Empresas Medianas

GainRatioAttributeEval		ChiSquaredAttributeEval		InfoGainAttributeEval	
Valor	Variable	Valor	Variable	Valor	Variable
0.5027	Registrado ChileProveedores	5.612,1579	Promedio Logeo Mensual	0.8758	Cantidad Ofertas
0.4891	Vigente ChileProveedores	5.557,1914	Cantidad Ofertas	0.8103	Cantidad de meses logeado
0.382	Cantidad Ofertas	5.299,9146	Cantidad de meses logeado	0.7839	Promedio Logeo Mensual
0.3653	Promedio Logeo Mensual	4.944,0709	Cantidad Adjudicaciones	0.7766	Cantidad Adjudicaciones
0.3512	Cantidad Adjudicaciones	2.641,8627	Ofertas Rechazadas	0.482	Vigente ChileProveedores
0.3082	Cantidad de meses logeado	2.585,5441	Tasa "Win rate"	0.4812	Tasa "Win rate"
0.2852	Estado Transaccional	2.573,3366	Cantidad Órdenes de Compra	0.4467	Monto Transado
0.2527	Ofertas Rechazadas	2.535,9359	Monto Transado	0.4379	Registrado ChileProveedores
0.1937	Tasa "Win rate"	2.535,3161	Vigente ChileProveedores	0.4162	Cantidad Órdenes de Compra
0.1811	Cantidad Órdenes de Compra	2.400,3937	Registrado ChileProveedores	0.3937	Ofertas Rechazadas
0.178	Monto Transado	1.256,3956	Estado Transaccional	0.1812	Tasa "Of-Rate"
0.1596	Cantidad de Reclamos	1.090,1915	Tasa "Of-Rate"	0.18	Estado Transaccional
0.1481	Tasa "Of-Rate"	376,0986	Cantidad de Reclamos	0.0712	Año Primera Interacción
0.046	Año Primera Interacción	375,9651	Año Primera Interacción	0.0595	Cantidad de Reclamos

(Fuente: Elaboración propia)

Anexo G.4: Empresas Grandes

Resultados de algoritmos de selección de atributos para las empresas tamaño grande. En negrita las variables finalmente seleccionadas.

Tabla 53: Selección de Atributos Empresas Grandes

GainRatioAttributeEval		ChiSquaredAttributeEval		InfoGainAttributeEval	
Valor	Variable	Valor	Variable	Valor	Variable
0.519	Cantidad Adjudicaciones	4.337,0314	Cantidad Adjudicaciones	1,1295	Cantidad Adjudicaciones
0.443	Cantidad Ofertas	3.430,7071	Cantidad Ofertas	0.9711	Cantidad Ofertas
0.44	Vigente ChileProveedores	3.384,0302	Monto Transado	0.8455	Monto Transado
0.428	Registrado ChileProveedores	3.203,0972	Cantidad Órdenes de Compra	0.796	Cantidad Órdenes de Compra
0.378	Cantidad Órdenes de Compra	2.320,4311	Cantidad de meses logeado	0.7463	Cantidad de meses logeado
0.336	Monto Transado	2.294,6180	Promedio Logeo Mensual	0.7099	Promedio Logeo Mensual
0.325	Cantidad de meses logeado	1.901,6373	Tasa "Win rate"	0.6566	Tasa "Win rate"
0.318	Promedio Logeo Mensual	1.386,3676	Ofertas Rechazadas	0.4682	Ofertas Rechazadas
0.264	Tasa "Win rate"	1.248,0453	Vigente ChileProveedores	0.422	Vigente ChileProveedores
0.255	Ofertas Rechazadas	1.113,8599	Registrado ChileProveedores	0.3554	Registrado ChileProveedores
0.225	Estado Transaccional	663,4558	Año Primera Interacción	0.1853	Año Primera Interacción
0.165	Cantidad de Reclamos	497,7685	Tasa "Of-Rate"	0.1471	Tasa "Of-Rate"
0.155	Año Primera Interacción	355,8895	Estado Transaccional	0.1216	Estado Transaccional
0.128	Tasa "Of-Rate"	240,4210	Cantidad de Reclamos	0.0742	Cantidad de Reclamos

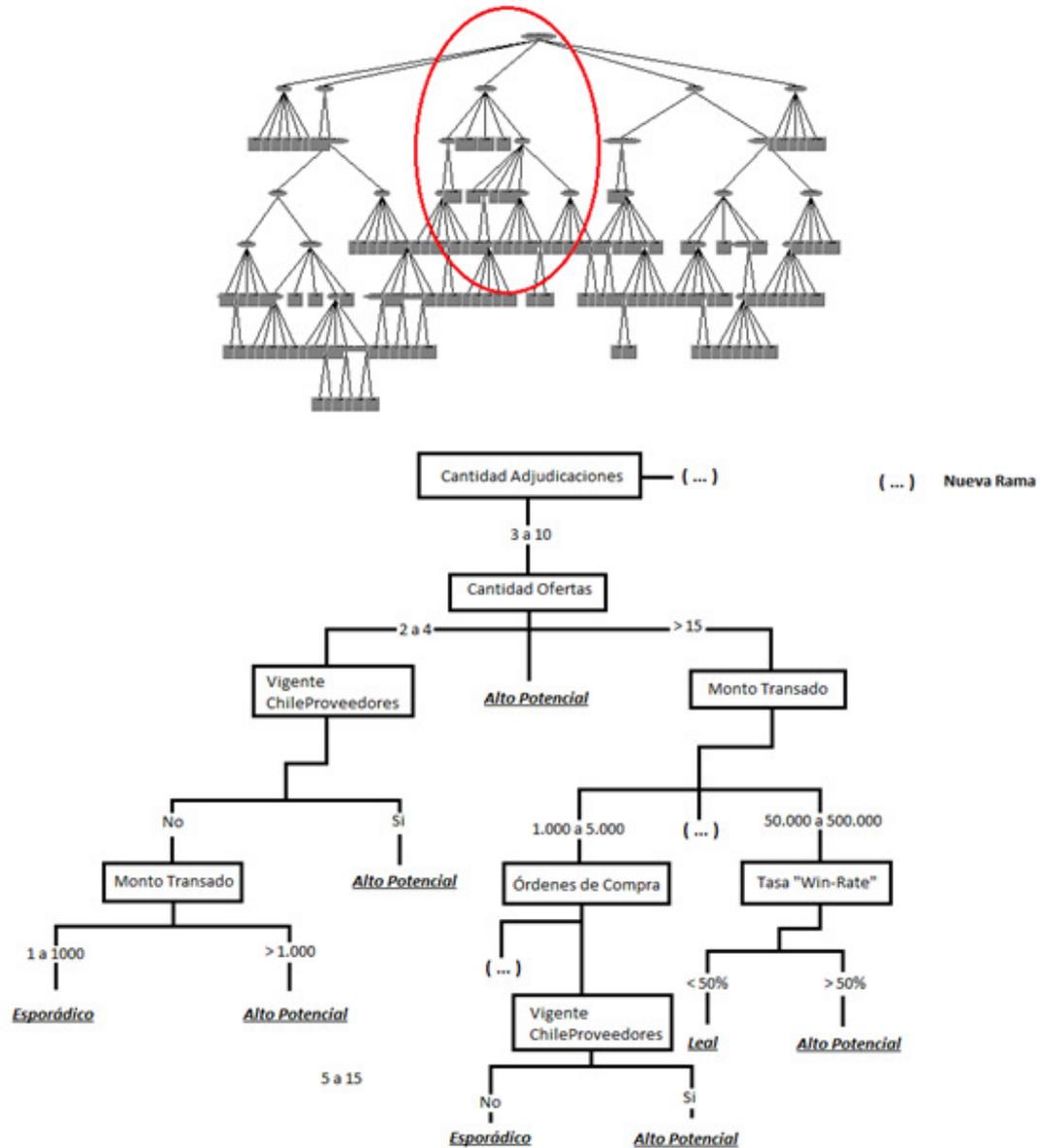
(Fuente: Elaboración propia)

9.8. Anexo H: Árboles de Decisión

Anexo H.1: Ejemplos de ramas

Mostrar el árbol de decisión solo tiene fines ilustrativos. Aquí se muestra el árbol de decisión de las grandes empresas (se muestra el de las grandes empresas por ser el más simple dentro de los 4 tamaños de empresa analizados).

Ilustración 38: Ejemplo de rama de árbol de clasificación



(Fuente: Elaboración Propia)

9.9. Anexo I: Reglas de Decisión

Anexo I.1: Ejemplos de reglas de Decisión

Lo siguiente corresponde a una serie de reglas de decisión creadas por el algoritmo PART. El objetivo de crearlas es comprender de mejor manera la clasificación

Al igual que en el ejemplo anterior de las ramas de árboles, se muestra un ejemplo del caso de las grandes empresas.

Regla 1: Cantidad Adjudicaciones = [>10] AND

Cantidad Ofertas = [>15] AND

Cantidad de meses logeado > 11 : Cluster leal

(Regla clasifica al 73% de los leales)

Regla 2: Cantidad Adjudicaciones = [0] AND

Monto Transado = [0]: Custer Candidatos a fuga

(Regla clasifica al 88% de los candidatos a fuga)

Regla 3: Cantidad de meses logeado ≤ 5 AND

Vigente ChileProveedores = NO AND

Cantidad Adjudicaciones = [1]: Cluster Esporádicos

(Regla clasifica al 70% de los esporádicos)

Regla 4: Cantidad Ofertas = [5 a 15] AND

Monto Transado = [5000 a 50000 USD]: Cluster Alto Potencial

(Regla clasifica al 30% de los proveedores de Alto Potencial)

Adicionalmente existen para el sector grande 80 reglas adicionales que van explicando los casos restantes. Estas 4 reglas (una para cada cluster) representan las variables más significativas por cluster. Así por ejemplo para los proveedores esporádicos la variable más relevante es la cantidad de meses logeado, que difiere de la variable relevante en el caso de los candidatos de alto potencial, quienes se caracterizan mejor analizando la cantidad de ofertas. En el caso de los otros dos clusters, la variable relevante es el nivel de adjudicaciones.