

UNIVERSIDAD DON BOSCO
VICERRECTORÍA DE ESTUDIOS DE POSTGRADO



TRABAJO DE GRADUACIÓN

MÉTODOS Y TÉCNICAS DE MINERÍA DE DATOS APLICADAS AL MARKETING
RELACIONAL BASADAS EN TECNOLOGÍA OPEN SOURCE.

PARA OPTAR AL GRADO DE:
MAESTRO EN ARQUITECTURA DEL SOFTWARE

ASESOR:
MSC. JOSÉ MARIO HERNÁNDEZ

PRESENTADO POR:
FRANCISCO RICARDO GUTIÉRREZ CÁCERES
FIDEL ASISCLO MENDOZA GUZMÁN

OCTUBRE DE 2015
ANTIGUO CUSCATLÁN, LA LIBERTAD, EL SALVADOR, CENTRO AMÉRICA

Métodos y Técnicas de Minería de Datos Aplicadas al Marketing Relacional Basadas en Tecnología Open Source

Francisco Ricardo Gutiérrez Cáceres, Fidel Asisclo Mendoza Guzmán

Centro de Estudios de Postgrados – Universidad Don Bosco

San Salvador, El Salvador

ricar_33@yahoo.com, asismg@gmail.com

Abstract – Este trabajo tiene por finalidad potenciar el uso de minería de datos en el sector comercio para las pequeñas y medianas empresas (PyME), haciendo uso de tecnologías Open Source. Proveyendo una guía de métodos, técnicas, herramientas y estrategias aplicadas a las actividades del marketing relacional: retención de clientes, rentabilidad por cliente y análisis de clientes potenciales, con el objetivo de causar interés en la investigación y apuesta tecnológica para la mejora de sus procesos de negocios, permitiéndoles ganar ventaja competitiva ante una globalización predominante para los que tienen el conocimiento de su mercado y nichos potenciales.

Palabras claves: Marketing relacional, minería de datos, perfil, rentabilidad, retención, segmentación de mercado.

I. INTRODUCCIÓN

1.1 La revolución de los mercados en el mundo han vuelto la actividad comercial compleja, la investigación de mercados se ha convertido en una necesidad para todo tipo de negocio. Las empresas grandes recurren a la estadística y a la acción minera en sus volúmenes de datos, en busca de estimaciones econométricas, demográficas y estudio de percepción de conductas de los clientes, previo a la presentación de un producto o servicio. La exploración de mercados en la actualidad es imprescindible ya que los competidores incrementan constantemente y los clientes son cada día más exigentes.

La gestión de marketing se basa en el esfuerzo consciente de las empresas para alcanzar un determinado nivel de intercambio con el público objetivo definido. La PyME ejecuta un marketing operativo que busca conquistar los mercados por medio de la acción a corto y medio plazo, centrado en el propósito de ventas en base a medios tácticos; desde esta perspectiva podría justificarse el poco interés por un marketing estratégico el cual pretende conocer las necesidades actuales y futuras de los clientes, así como nuevos nichos, segmentación, interés y valor potencial de los mercados ya que a través del

marketing operativo tienen un contacto directo con sus clientes y ejecutan lo necesario en el momento preciso.

Sin embargo el marketing estratégico se orienta a satisfacer aquellas necesidades del consumidor que constituyen oportunidades de negocio atractivas para las empresas, orientando las actividades a través de la formulación de objetivos que considere toda la organización enfocada al mercado y así crear ventajas competitivas, indicando a la empresa qué acciones debe realizar de acuerdo a su naturaleza de tal manera orientándolas a las oportunidades de rentabilidad que presenten sus productos o servicios en determinado tiempo y segmento de mercado.

Por tal razón el cometido de este artículo es servir como guía inicial para aquellas empresas del sector PyME que tengan el deseo utilizar minería de datos para la investigación de mercados.

La secuencia del documento presenta la minería conceptual y su estructura algorítmica clasificada en métodos y técnicas; describe tres actividades del marketing que son de interés para toda organización; también se abordan las herramientas que pueden utilizarse para la aplicación de estas actividades en explotación minera; además incluye la propuesta de una tecnología colaborativa para las gestiones de las relaciones con el cliente. Concluye con la aplicación de casos específicos a medida de ejemplo utilizando software Open Source."

El documento aporta utilidad para todas aquellas personas interesadas en la investigación de mercados orientada a las mejoras de las relaciones con los clientes, posicionamiento de mercado, esfuerzos focalizados y una visión estratégica competitiva.

II. MÉTODOS Y TÉCNICAS DE MINERÍA

2.1.1 ¿En qué consiste la minería de datos?

Minería de datos es un concepto tan evolucionado hoy en día que algunos autores hacen aproximaciones por fases, incluyendo los elementos claves y más utilizados por los expertos, a fin de denotar importancia e impacto que tiene la

aplicación del análisis de la información en diferentes disciplinas. A continuación se tienen las siguientes interpretaciones.

“La minería de datos puede definirse inicialmente como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos” (César Pérez, 2008, p. 1)

Así también otro autor señala que:

“La minería de datos, es un proceso de análisis de datos de naturaleza interdisciplinaria, cuya propuesta es la identificación y extracción de conocimientos de alto valor a partir de los datos. Los conjuntos de datos pueden ser de alto o bajo volumen, tener muchos atributos descriptivos, estructuras poco claras, e incorporar datos desconocidos, erróneos y con ruido” (Nettleton David, 2005, p. 14)

Uno de los objetivos principales de la minería de datos es emplear el valor de la información encontrada a través de los patrones, obteniendo con ello un mayor conocimiento del negocio para toma de decisiones acertadas.

«Por patrones se entiende las relaciones de comportamiento no visibles de los clientes en la adquisición de un producto o servicio»

Los algoritmos aplicados en la minería tienen implícitas tareas descriptivas, predictivas, clasificación y ajustes de modelos que permiten presagiar comportamiento de las variables entre otros.

La minería consiste en la profundización, extracción y transformación de los datos en información, que residen de manera explícita en los almacenes de datos.

Minando se da un significado a los datos, convirtiéndose en conocimiento útil para en quienes recae la responsabilidad de tomar decisiones en los diferentes sectores. La minería es un área multidisciplinar que tiene sus bases en la estadística por medio del razonamiento matemático y la aplicación de sus principios. Apoyándose en tecnologías avanzadas actualmente se convierte en un proceso de negocio, que permiten entre muchas cosas la optimización de recursos, ventaja competitiva, mejora de procesos y relaciones con los clientes entre otras.

2.1.2 ¿Cuál es el objetivo de las técnicas de minería?

La minería utiliza diferentes técnicas para el análisis y procesamiento de los datos, orientadas al descubrimiento de conocimiento y generación de modelos. “Estas técnicas tienen por objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas de análisis de datos” [1]. Su clasificación contempla las técnicas predictivas,

en las que las variables se pueden catalogar inicialmente en dependientes e independientes; así como un análisis multivariante. Además incluye técnicas descriptivas en las que todas las variables tienen el mismo estatus con una perspectiva de interdependencia, y por último las técnicas auxiliares.

2.1.3 Técnicas predictivas

Especifican el modelo para los datos en base a un conocimiento teórico previo. La aplicación del modelo se deberá verificar antes de aceptarlo como válido considerando las fases de:

- a) *Identificación objetiva:* aplicación de reglas que permitan identificar el mejor modelo posible que se ajuste los datos.
- b) *Estimación:* cálculo de los parámetros elegidos.
- c) *Diagnosis:* validez del modelo estimado.
- d) *Predicción:* proceso de utilización del modelo identificado, estimado y validado para la predicción de valores futuros de las variables dependientes.

2.1.4 Técnicas descriptivas

No se asigna un papel predeterminado a las variables, no supone la existencia de variables dependientes e independientes y ningún modelo previo para los datos. Los modelos se crean automáticamente partiendo del reconocimiento de patrones.

Tanto las técnicas predictivas como las descriptivas están enfocadas al descubrimiento del conocimiento embebido en los datos.

2.1.5 Técnicas auxiliares

Se trata de nuevos métodos basados en técnicas estadísticas descriptivas, consultas e informes dirigidos principalmente a la verificación.

La siguiente figura muestra una clasificación de métodos y técnicas de minería de datos.

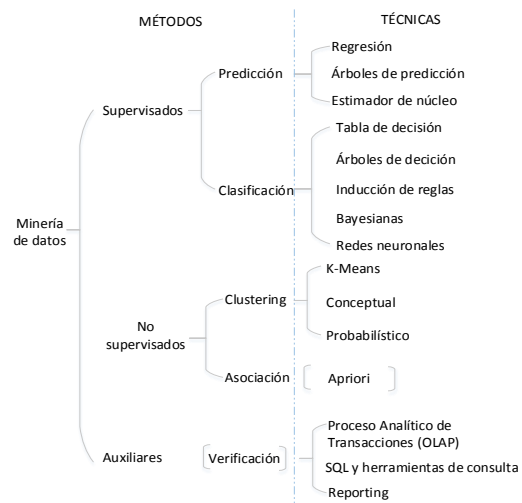


Figura 1. Clasificación de los métodos y técnicas de minería de datos.

En relación metodológica para la aplicación de las técnicas descritas se definen los métodos de mayor importancia acorde a la finalidad de éste artículo:

- a) *Clasificación*: consiste en clasificar los objetos, entre un rango de categorías por ejemplo mapas asociativos de memoria, entre los que se encuentran las redes de Kohonen, redes neuronales que clasifican entes mediante aprendizaje por esfuerzo.
- b) *Clustering*: agrupan los objetos según su similitud, formando conjuntos o clases. La diferencia con la clasificación es que no hay división previa del espacio en categorías o clases.
- c) *Asociación*: el objetivo es el de descubrir relaciones ocultas entre los objetos, o incluso entre los propios atributos de los objetos.
- d) *Predicción*: modelos predictivos con el fin de validar hipótesis de comportamiento preconcebidas.

2.2.1 ¿Cuál es la metodología de aplicación de técnicas de minería datos?

Las diferentes técnicas posibles de la minería pueden aplicarse mediante distintos métodos teniendo cada uno sus propias ventajas y desventajas, en muchas ocasiones se utilizan combinaciones de varios métodos para alcanzar un solo objetivo. Algunos de estos métodos se describen a continuación.

2.2.2 Métodos supervisados

Predicen el valor de un atributo “etiqueta” de un conjunto de datos, estas relaciones sirven para realizar la predicción cuya etiqueta es desconocida. Esta forma de trabajar se conoce como aprendizaje supervisado y se desarrolla en dos fases:

- a) *Entrenamiento*: construcción de un modelo usando un subconjunto de datos con etiqueta conocida.
- b) *Prueba*: prueba del modelo sobre el resto de los datos. En este grupo se encuentran algoritmos que resuelven problemas de clasificación debido a que trabajan con etiquetas discretas (árboles de decisión, tablas de decisión e inducción neuronal) y algoritmos que se utilizan en la predicción de valores continuos como son la regresión o las series temporales.

2.2.3 Métodos no supervisados

Algoritmos no supervisados realizan tareas descriptivas como el descubrimiento de patrones y tendencias en los datos actuales (no utilizan datos históricos), esa información sirve para llevar a cabo acciones y obtener un beneficio científico o de negocio. La inducción de reglas de asociación es la técnica más utilizada de este grupo.

2.3.1 ¿Por qué usar minería de datos?

Las razones pueden ser muchas por la universalidad aplicada de la minería, sin embargo en el marketing relacional permite afrontar mejor el actual ambiente competitivo.

El uso de minería potencia en el sector comercio el análisis de los indicadores siguientes:

- Identificación de oportunidades del negocio.
- Identificación del tamaño y composición del mercado.
- Perspectivas de crecimiento del mercado.
- Elementos y características del mercado objetivo.
- Análisis de clientes potenciales.
- Comportamiento y estrategias de la competencia.
- Diseño y adaptación del producto.
- Patrones de consumo de los productos o servicios.
- Pronóstico de la demanda.
- Entre otros.

Además de la obtención de los indicadores descritos anteriormente, la minería podrá ayudar a la PyME a controlar, optimizar, administrar, planificar, predecir, negociar y tomar decisiones fundadas en el descubrimiento de conocimiento a través de la aplicación de técnicas de mineras. Sin embargo pese a que los beneficios pueden ser innumerables, la PyME está limitada en tiempo y presupuesto, para optar en invertir en tecnología y recursos, para la constante investigación en la obtención de esos valiosos indicadores. De acuerdo a la información que se muestra en la *Tabla 1*, se puede ver la penetración de mercado y el nivel tecnológico que sostienen los procesos de negocio de la PyME.

Tamaño	Empleados	Activo fijo	Organización	Tecnología	Participación en el mercado
Micro	De 0 a 10		Informal	Artesanal	Consumo final
Pequeña	De 11 a 20	Hasta \$85,714.28	Dos niveles	Rudimentaria	Local
Mediana	De 21 a 99	Hasta \$228,571.42	Tres niveles	Bajo nivel tecnológico	Nacional o regional
Grande	Más de 100	Mayor de \$228,571.42	Más de tres niveles	Altamente sofisticado	Mundial

Tabla 1: Clasificación de las empresas Salvadoreñas según Banco Central de Reserva de El Salvador

Podríamos preguntarnos ¿Es por falta de capital el rudimentario o bajo nivel tecnológico? ¿Poca importancia a la apuesta tecnológica? o ¿Desconocimiento de los pasos a seguir? Independientemente, como se mencionó al inicio, animamos a la PyME embarcarse en las posibilidades que brinda la minería a través del uso de software libre, métodos y elementos necesarios que damos a conocer para aprovechar de la mejor manera la información que nos brindan los clientes. Bastará con interesarse, tomar un puñado de variables claves, definir las actividades, aplicar los métodos que les proponemos y analizar los resultados que buscamos.

2.3.2 ¿Dónde se utiliza la minería de datos?

Puntualmente en nuestro caso en el marketing relacional. En aquellas particularidades estratégicas del negocio que sean vitales para tomar ventaja competitiva y permitan conocer mejor a mis clientes, además cómo los procesos que actualmente sigo para dar un producto o servicio son los más

cercanos a lo que ellos buscan y que a su vez generan mejores réditos.

Técnicamente podemos decir que la minería es utilizada en: la gestión de análisis de datos para la mejora del producto o servicio, enfocado a gestión de clientes (base del mercadeo relacional).

III. ACTIVIDADES DEL MARKETING RELACIONAL

3.1.1 ¿Qué es el marketing relacional?

Un enfoque relacional dirigido al cliente, que gana aceptación y trasciende a través de la evolución de las tecnologías de información, de lo comercial hasta lo estratégico empresarial, cubrir todo lo que involucra en una definición es desafiante, sin embargo algunos autores proponen:

“El marketing relacional consiste en atraer, mantener e intensificar las relaciones con los clientes” (Berry, 1983).

“Proceso social y directivo de establecer y cultivar relaciones con los clientes, creando vínculos con beneficios para cada una de las partes, incluyendo a vendedores, prescriptores, distribuidores y cada uno de los interlocutores fundamentales para el mantenimiento y explotación de la relación” (Alep, 1994, p. 35).

Uno de los objetivos claves del marketing actual es establecer relaciones firmes y duraderas con las personas o con las organizaciones que directa o indirectamente podrían influir en el éxito de las actividades de marketing de la empresa. El marketing relacional tiene por objeto establecer relaciones mutuamente satisfactorias y de largo plazo, con los participantes claves (consumidores, proveedores, distribuidores y otros socios) con el fin de conservar e incrementar el negocio. Permitiendo así crear fuerte vínculos económicos, técnicos y sociales entre las distintas partes.

El enfoque del marketing relacional significó una filosofía que en lugar de seguir centrada en el producto, “fabricar y vender” se cambió a “detectar y responder”. En lugar de “cazar” se empezó a “cultivar”. El marketing no se concentraba en encontrar al consumidor adecuado para el producto, sino en desarrollar los productos adecuados para los consumidores.

Para que una empresa pueda establecer relaciones sólidas necesita conocer las capacidades y los recursos de diferentes grupos, así como sus necesidades, objetivos y deseos. Gran parte de las empresas perfilan ofertas, servicios y mensajes independientes para los distintos consumidores, recopilando información sobre las transacciones de cada consumidor, información demográfica, psicográficas sobre sus preferencias de distribución y medios de comunicación.

“La función de cualquier empresa es ofrecer valor a sus clientes a cambio de utilidades, y en una economía globalizada; un número creciente de compradores ante un

abanico de ofertas, decidirán por aquella empresa que agudice el proceso de valor, seleccionando, ofreciendo y comunicando un importe superior a sus clientes” (Kotler Philip, Keller Lane Kevin, 2006, p. 36).

Las actividades del marketing son numerosas así como sus beneficios; por mencionar unas tenemos la siguiente imagen:



Figura 2: Actividades del marketing relacional.
Fuente: Elaboración propia.

Los esfuerzos mineros de éste trabajo se centran en tres actividades del marketing relacional, con finalidad de despertar el interés de las PyME y apostar a estas tecnologías para beneficiarse, dichas actividades son las siguientes:

- Retención de clientes.
- Rentabilidad por cliente.
- Análisis de clientes potenciales.

Seguidamente se abordan las características e importancia de cada una en un contexto de mercado.

3.1.2 Retención de clientes

La retención de clientes implica dedicar menos tiempo a buscar otros y cuidar más a los existentes. La retención está centrada en ofrecer lo que la gente quiere y necesita.

Retener los clientes no solo resulta bueno para los intereses de la organización, los clientes también se benefician de la creación de relaciones con las personas que les proporcionan los bienes y servicios. Aunque conscientemente sabemos que ambas partes pueden sacar provecho es indispensable plantear:

¿Por qué es importante la retención de un cliente?

Conservar los clientes puede tener más de una perspectiva, en el contexto cualitativo la pérdida de un cliente podría afectar la imagen de la empresa, pues un cliente insatisfecho decidirá no volver y con alta probabilidad trasladará su insatisfacción a otros clientes potenciales, generando resultados nocivos al prestigio de la empresa como raíz de las malas experiencias en las relaciones de compra venta con la organización.

En lo cuantitativo el análisis está orientado a conocer el valor en términos de rentabilidad del cliente, contra puesto con lo que significará perderlo o el costo que representará el tratar de retenerlo. Sabemos que todo cliente es importante; pero como ente directivo en las empresas es requisito medir los esfuerzos frente a los resultados esperados. Bajo este enfoque hay que tener en claro que el perder un cliente no necesariamente significará una pérdida, por el contrario podría ser hasta ganancioso perder esa relación.

Sin embargo esas decisiones solo se podrán tomar con la plena seguridad del impacto que la organización tendrá al invertir en la retención o ruptura de esos lazos. Y para ello la minería es clave, por medio de la segmentación el gerente conoce con exactitud a sus clientes, sus gustos, gasto promedio, frecuencia de compra, categorías, canal, edades, sexo, distancias, entre otros. Esto nos llevará a ¿Qué tipos de clientes tenemos y qué estrategia podemos seguir con cada uno? ¿A quiénes retener y por qué?

Los consultores internacionales de Bain & Company han sugerido que: *“puede llegar a ser hasta cinco veces más costoso captar un cliente nuevo que retener uno actual”*.

¿Entonces por qué un cliente retenido debería ser más rentable que uno nuevo?

De acuerdo con Reichheld y Sasser (1990) existen varias razones:

En primer lugar, el coste de adquirir nuevas cuentas puede ser significativo, y requiere tiempo, incluso años, convertir a un nuevo cliente en uno rentable.

En segundo lugar, los clientes que están más satisfechos con la relación son los más propensos a realizar una mayor proporción de sus compras con ese proveedor, incluso hasta el extremo de ser este una fuente única.

En tercer lugar, a medida que la relación evoluciona, hay una mayor colaboración y comprensión mutua que sirve para reducir los costes. Es más fácil vender a los clientes retenidos, y las economías de escala producen costes operativos más bajos. Estos clientes también están más dispuestos a integrar sus sistemas de tecnología de la información (por ejemplo, sus sistemas de planificación, programación y pedidos) con el proveedor lo cual conduce a otras reducciones del coste.

En cuarto lugar, los clientes satisfechos suelen recomendar la compañía a otros, lo cual promueve la generación de beneficios, ya que el coste de adquirir a estos nuevos clientes es mucho más reducido. Finalmente, los clientes leales son a menudo menos sensibles al precio y menos propensos a cambiar de proveedores por el aumento de precios.

Además según Wilfredo Pareto ha demostrado que el 70 u 80% de las ventas se concentran en el 20 o 30% de los clientes que serían precisamente a quienes se deben de retener. Sin embargo retener no es un acto solo de cuantificar o masificar grupos, sino valorar estratégicamente el impacto a la organización en todo contexto sobre el soltar o mantener la relación.

3.1.3 Rentabilidad de clientes

La rentabilidad de un cliente puede medirse a través del valor, frecuencia y lealtad, considerándose todos los flujos de ingresos menos los costos de producto y servicio, costos de adquisición y marketing. “La medición de la rentabilidad del cliente supone la estimación de ingresos y costes asociados a cada uno y como resultado la rentabilidad individualizada” [4]. Permite visualizar los beneficios generados y estimados a lo largo del tiempo en la relación con la organización. La retención y rentabilidad están fuerte y estrechamente relacionadas, pues la predicción de deserción de un cliente es tan importante y más cuando su rentabilidad es alta para la empresa, la identificación de estas tendencias a futuro permiten focalizar los esfuerzos para solidificar las relaciones. La figura 3 muestra el ciclo de retención de cliente.



Figura 3: Ciclo de retención de clientes.
Fuente: Elaboración propia.

¿En qué consiste la rentabilidad de los clientes?

La rentabilidad es una relación entre los recursos necesarios y el beneficio económico que deriva de ellos. Es considerada como ROE (“return on equity”, rentabilidad sobre recursos propios). Algunos autores proponen las siguientes definiciones.

“La rentabilidad es una medida que relaciona los rendimientos de la empresa con las ventas, los activos o capital. Esta medida permite evaluar las ganancias de la empresa respecto a un nivel dado de ventas, de activos o de inversión de los dueños” (Guitman, Lawrence J, 1992, p. 62 y 171).

“La rentabilidad mide la eficiencia general de la gerencia, demostradas a través de las utilidades obtenidas

de las ventas y por el manejo adecuado de los recursos, es decir la inversión de la empresa” (Guiltinan, 1984, p. 117).

“Desde el punto de vista de la inversión de capital la rentabilidad es la tasa mínima de ganancia que una persona o institución tiene en mente, sobre el monto de capital invertido en una empresa o proyecto” (Baca Urbina, 1987, p. 79).

En esencia podemos decir que la rentabilidad es el porcentaje o tasa de ganancia obtenida por la inversión de un capital determinado. La figura 4 muestra el cálculo de la rentabilidad.

$$\text{Rentabilidad (R)} = \frac{\text{ingresos} - \text{inversión}}{\text{inversión}} \text{ o, } R = \frac{\text{beneficios}}{\text{inversión}} \text{ o, } R = \frac{\text{ingresos}}{\text{inversión}} - 1$$

Figura 4: Cálculo de rentabilidad.

Fuente: Van Horne, James C. Financial Management and Policy (1995).

Guiltinan propone que el análisis de rentabilidad se debe hacer a partir de la diferencia resultante entre los ingresos por ventas y los costos variables del producto, en el entendido de que esta diferencia es la parte que aporta el producto, como unidad de negocio, para contribuir a cubrir los costos fijos directos controlables, si los hubiera, los costos fijos indirectos (gastos generales, administrativos, impuestos, entre otros.) y las utilidades.

“La rentabilidad tiene un efecto producente cuando agregamos la variable *valor del cliente* centrado en la satisfacción de sus necesidades y no las del negocio, definiéndose como la suma de los beneficios percibidos al comprar y utilizar un producto o servicio” [14].

“El objeto de estudio del marketing se centra en las relaciones de intercambio de valor con sus clientes. Si una organización quiere incrementar su valor debe identificar las necesidades de sus clientes y gestionar estas relaciones de manera que cree, comunique y entregue propuestas de valor que satisfagan a todas las partes (stakeholders). Más en concreto, la gestión del valor del cliente implica la aproximación al cliente de forma individualizada y con una orientación a largo plazo, de manera que se capten y retengan clientes rentables, ofreciendo productos y/o servicios nuevos (cross-selling) y más rentables (up-selling), prolongando a su vez el tiempo de relación con la organización” (Esteban et al., 2008).

Esta estrategia relacional busca hacer una oferta a cada cliente (o pequeños grupos de clientes) que maximice el valor para el cliente (funcional, social, emocional, epistemológico y condicional) y el valor del cliente para la

empresa (en términos de tasa de éxito, ingresos, coste y duración). En este sentido, los beneficios para la empresa al mantener altos niveles de retención de clientes rentables se concretan en el mantenimiento de un crecimiento vegetativo muy importante, en la creación de barreras de entrada a los competidores, en la disposición de una amplia base de clientes para nuevos productos y en la consolidación del modelo de calidad de servicios¹.

La siguiente figura muestra el valor presente neto en función de rentabilidad, relación, costos y ventas generadas.

$$\text{Rentabilidad (Margen de Contribución Variable)} = \frac{\text{Utilidades (contribución variable)}}{\text{Ventas}}$$

Figura 5: Cálculo de rentabilidad.

Fuente: Gitman, Lawrence J. "Administración Financiera Básica (1992).

Sigue siendo muy evidente hasta el momento del por qué se requiere aplicar minería de datos en las organizaciones y más aquellas que se encuentran en crecimiento constante. La PyME puede sacarle provecho apostándole a tecnologías Open Source e iniciando con estas actividades del marketing relacional propuestas en este trabajo.

A continuación se aborda la actividad de captación de clientes.

3.1.4 Clientes potenciales

La captación de clientes tiene implícita la segmentación con características homogéneas basadas en las necesidades, identificado todos los segmentos posibles. “La gran mayoría de las empresas están inmersas en tácticas de captación masiva e indiscriminada de perfil cortoplacista” [12]. Sin embargo no se trata de captar clientes a cualquier precio, sino estratégicamente para conocer quien se incorpora a nuestra empresa, qué es capaz de aportarnos, cuánto tiempo puede permanecer con nosotros y qué puede generar durante ese tiempo. Toda estrategia de captación debe tener como objetivo principal la rentabilidad, definiendo cuanto invertir en cada segmento de cliente, qué canales utilizar y qué recursos comerciales de marketing. Para captar de forma inteligente, las empresas deben establecer planes de captación selectiva en los que el eje sea el valor de cada segmento².

¿Quiénes son los clientes potenciales?

En primer lugar hay que tener en claro que se está buscando y dónde lo estamos buscando. Para ello se debe hacer una descripción y crear un perfil para identificarlo. Algunos de esos elementos pueden ser:

- Rango de edades.

¹ Managing Customers for Profit: Strategies to Increase Profits and Build Loyalty (2008).

² Captación de clientes: claves para diseñar Estrategias Rentables de Adquisición y Prospección Selectiva. Mombiedro, Rafael - Valdés, Pedro.

- Nivel de ingreso.
- Estado civil.
- Sexo.
- Nivel educativo.
- Entre otros.

La segmentación es inherente en la captación de clientes potenciales. Kotler propone el uso de las cuatro “P” para un marketing estratégico:

- *Investigación (Probing): es la que nos lleva a qué clientes queremos dirigirnos.*
- *Segmentación (Partitioning): técnica que apunta a un grupo de clientes con características similares.*
- *Clasificación (Prioritizing): una vez hecha la segmentación de los clientes, priorizamos a qué segmentos nos vamos a dirigir, por ventaja competitiva o simplemente porque nos interesa dirigirnos a ellos.*
- *Posicionamiento (Positioning): a cada uno de los segmentos a los que se decida ir, se debe definir con precisión los puntos fuertes que se pueden aportar para diferenciarnos y posicionarnos.*

Cross selling: Se basa en el conocimiento del perfil del cliente. Cuando éste se determina se compara su cesta de compras con la que mayoritariamente un cliente óptimo ideal de su tipología haría y de esta manera se le ofrecen productos de los que la empresa comercializa que aún no compra pero que teóricamente por pertenecer a ese segmento debería comprar. Esta técnica está basada indispensablemente en el conocimiento del cliente y *el control exhaustivo de lo que una masa global de clientes compra*, y posteriormente segmentados por tipologías.

Por un lado el cliente aumentará su consumo, con lo que contribuye a un mayor margen para la empresa. Por otro lado, al ya ser un cliente existente, el coste de captarlo es casi nulo, con lo cual se incrementará notablemente el beneficio neto al tener menores gastos comerciales. Además un factor importante en el desarrollo de una alta calidad en esta relación es que el cliente nunca es bombardeado con una infinidad de productos que no sean afines a él.

IV. HERRAMIENTAS DE MINERÍA

Entenderemos por herramientas los instrumentos que permitan realizar un determinado trabajo, efectuándolo con facilidad y prontitud.

El trabajo de minería de datos sería imposible realizarlo sin herramientas informáticas, aunque un algoritmo es la sucesión ordenada de operaciones para la realización de un cálculo y el resultado de este se obtiene siguiendo sistemáticamente sus pasos, sería insostenible para un equipo de personas llevar a cabo estos procedimientos de forma manual.

En la actualidad existen muchas herramientas sofisticadas para el ejercicio minero de la información con una complejidad de oportunidades para el análisis y explotación de los datos incalculables tanto propietarias como Open Source. Sin embargo como se planteó al inicio en este trabajo nos limitaremos a abordar software libre con la finalidad de dar a conocer a la PyME la existencia de las mismas, el nivel de madurez y oportunidad que representan para la obtención de tantos beneficios en la implementación estratégica del marketing relacional. A continuación se describen las más reconocidas así como sus enlaces de consulta y descarga en internet.



WEKA (Waikato Environment for Knowledge Analysis): Escrito en lenguaje Java, suite de software para el análisis y descubrimiento de conocimiento con soporte para varias tareas como: proceso previo, agrupamiento, clasificación, regresión, visualización y características de selección. Sus técnicas se basan en la hipótesis de que los datos están disponibles en un único archivo plano o relación, donde cada punto marcado es etiquetado por un número fijo de atributos. Proporciona acceso a base de datos SQL. Su interfaz de usuario puede ser gráfica o a través de líneas de comando. Actualmente está la versión estable 3.7.5 lanzada el 31 de octubre de 2011.

Para consulta de requerimientos, documentación y descarga de la herramienta visite el sitio web:

[<http://www.cs.waikato.ac.nz/~ml/index.html>]



ORANGE: permite una programación visual, rápida y versátil para un análisis exploratorio de datos, interfaz amigable e intuitiva. Entre sus características de minería permite pre-procesamiento, filtros, modelación, exploración y evaluación de modelos. Sus componentes algorítmicos están desarrollados en C++ y su interfaz gráfica en Python. Se encuentra disponible en el sitio web:

[<http://orange.biolab.si/>] última versión estable 2.7 lanzada el 8 mayo de 2015.



KNIME (Konstanz Information Miner): Permite crear visualmente flujos de datos, ejecutar análisis selectivamente, modelar y generar vistas interactivas. Está disponible en varias distribuciones siendo una de las más completas KNIME Server.

KNIME está desarrollado sobre la plataforma Eclipse y programado esencialmente en java. Está concebido como una herramienta gráfica y dispone de una serie de nodos (que encapsulan distintos tipos de algoritmos) y flechas (que representan flujos de datos) que se despliegan y combinan de manera gráfica e interactiva.

Los nodos implementan distintos tipos de acciones que pueden ejecutarse sobre una tabla de datos:

- Manipulación de filas, columnas, muestreos, agrupaciones y transformaciones.
- Visualización (histogramas).
- Creación de modelos estadísticos y de minería de datos, como árboles de decisión, máquinas de vector soporte y regresiones.
- Validación de modelos, como curvas ROC.
- Scoring o aplicación de dichos modelos sobre conjuntos nuevos de datos.

KNIME integra diversos componentes para aprendizaje automático y minería de datos a través de su concepto de fraccionamiento de datos (data pipelining) modular. La interfaz gráfica de usuario permite el montaje fácil y rápido de nodos para pre-procesamiento de datos (ETL: extracción, carga, transformación), para el análisis de datos, modelado y visualización. KNIME es desde 2006 utilizado en la investigación farmacéutica, pero también se utiliza en otras áreas, como: análisis de datos de cliente de CRM, inteligencia de negocio y análisis de datos financieros³. Su descarga y documentación puede hacerse a través del sitio web: [http://www.knime.org/]



RAPID MINER: (YALE = Yet another learning environment): es un entorno para

aprendizaje mecánico y experimentos de data mining, utilizado tanto en investigación como en tareas de día a día por diferentes empresas. Produce sus resultados en archivos XML y cuentan con la interface gráfica del mismo programa. Provee más de 500 operadores para los principales procesos de aprendizaje en máquina y al tiempo combina esquemas y atributos de evaluación.

La herramienta puede ser utilizada para los sectores de industria siguientes:

- Automotores.
- Banca.
- Seguros.
- Ciencias naturales.
- Fabricación.
- Petróleo / Gas.
- Consumo y comercio.
- Telecomunicaciones.

Aplicada a casos de uso como:

- Prevención de abandono.
- Valor de vida del cliente.
- Segmentación de clientes.
- Mantenimiento predictivo.

- Propensión de productos.
- Seguro de calidad.
- Modelado de riesgos.
- Análisis de sentimientos.
- Ventas cruzadas.

De igual manera que las anteriores esta herramienta está disponible en el sitio: [https://rapidminer.com/]

Existen más herramientas bajo licenciamiento GNU-GPL, sin embargo se presentan las más conocidas, robustas, en constante desarrollo, soporte vigente y sobretudo multiplataforma.

V. TECNOLOGÍA CRM ESTRATÉGICA CENTRADA EN EL CLIENTE

Según Lee Iacocca (2002), la compañía que tenga el mejor sistema de distribución y el mejor servicio, tiene todas las de ganar.

“El marketing de transacciones se centra en la transmisión de mensajes al cliente, y el marketing de relaciones pretende iniciar y mantener diálogos con el cliente. Recibir información de los clientes es tanto o más importante que transmitir información. Esto tiene grandes implicaciones para el negocio de una empresa” (Pinto, 1997, p. 10).

Algunos autores como Pitta y Peppers (1998) proponen que para el establecimiento de un marketing “one to one” deben existir cuatro pasos claves:

- Identificación de los clientes.
- Diferenciación de los clientes.
- Interacción con cada cliente.
- Ofrecimiento de un producto o servicio personalizado.

“CRM es una estrategia de negocio que permite anticiparse a la satisfacción de las necesidades del cliente, con base en el conocimiento de sus gustos y prioridades, que además soporta una arquitectura integrada de procesos, tecnología y cultura de la organización” [15].

Ignacio García Valcárcel (2001) Estableció que el CRM es una estrategia de negocio porque engloba a toda la organización ya que requiere en gran parte de las implantaciones, un rediseño del negocio donde el cliente participa en forma directa en el modelado de la empresa, teniendo en cuenta sus necesidades.

CRM es la estrategia de negocio soportada por una herramienta tecnológica como es el software, por la cual una empresa pretende apalancar en el conocimiento de sus clientes para crear relaciones más rentables basadas en la aportación de

³ Wikipedia (2015)
https://es.wikipedia.org/wiki/KNIME

valor añadido a sus productos o servicios que finalmente beneficien a sus clientes⁴.

5.1.1 ¿La aplicación de un CRM ayuda a mejorar la relación con el cliente?

“El objetivo de un CRM es construir relaciones duraderas mediante la comprensión de las necesidades y preferencias individuales, de este modo añade valor a la empresa y al cliente.

Esta estrategia produce un proceso de evolución y adaptación de un todo hacia lo particular, de lo masivo a lo individual, clasifica al consumidor dentro de un segmento.

Una metodología CRM trata de utilizar la tecnología, para organizar, automatizar y sincronizar los procesos de negocio, principalmente actividades de ventas, comercialización, servicio al cliente y soporte técnico entre sus cometidos se tienen” [9]:

- Encontrar, atraer y ganar nuevos clientes.
- Retener clientes actuales en la empresa.
- Atraer a antiguos clientes.
- Reducir costos de marketing y servicio al cliente.

Además una tecnología CRM ayuda a conocer todos los diferentes puntos de contacto con los cuales el cliente interactúa con la empresa. Esto permite ser más certero en la determinación del canal como medio de contacto directo con el cliente.

Un proceso CRM determina cuatro pasos dentro de una estrategia competitiva orientada al cliente con el rediseño de los procesos de negocio encaminado a satisfacer al cliente y una cultura de empleados como responsables últimos de la atención al cliente, todo esto con la tecnología CRM dando soporte integral.

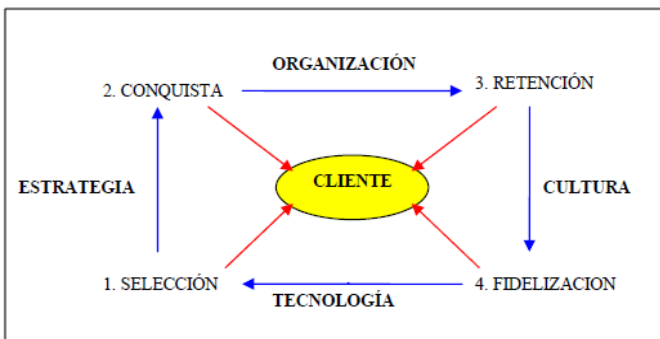


Figura 6: Proceso de fidelización CRM según (Cuesta 2003).

Un CRM, ha sido normalmente visto por las empresas como una actividad relacionada a desarrollar y retener a los clientes a través de aumentar su satisfacción y lealtad. CRM es un proceso diseñado para recoger información de los clientes, comprender

sus características y aplicar éstas en actividades específicas del marketing.

El éxito de una implantación de CRM en una empresa está en asegurar que la organización de la empresa está alineada con la estrategia de sus clientes. Las razones por las que las tecnologías CRM fallan en su implantación son por pensar que la solución de software resolverá los problemas de relaciones con sus clientes. CRM no es solo un paquete de software; implica una completa reconsideración de organizarse dentro de la compañía a nivel de procesos. Una vez se defina la estrategia que se desee llevar con los clientes y la empresa sostenga dicha táctica, sería el momento de implantar la herramienta colaborativa CRM.

Algunas de las razones por la cuales se implantaría un CRM serían:

- Mejora de nivel de satisfacción del cliente.
- Retención de los clientes existentes.
- Mejora en el valor en tiempo del cliente.
- Mejora en estrategias de ventas, marketing y finanzas.
- Atracción de nuevos clientes.
- Ahorro en costos de campañas.

Poner en marcha una estrategia CRM implica más que simplemente crear soluciones que den soporte a áreas de gestión de información de clientes (como la automatización del marketing), sino más bien es que estas iniciativas obedecen a una serie de principios que hacen que el negocio evolucione alrededor del conocimiento de clientes. Una diferencia de esta evolución es, por ejemplo, pasar de crear productos en cantidades masivas a bajo coste, a crear productos o servicios más personalizados que se acoplan a uno o varios segmentos de mercado específicos. Esta evolución solamente es posible si en realidad se conoce al cliente y si la empresa es capaz de reaccionar con una propuesta de valor⁵.

Se puede decir que entre los objetivos principales de una empresa las prioridades más altas en referencia a su estrategia comercial incluyen, en orden de importancia:

- Obtener nuevos clientes (aumentar la cuota de mercado).
- Aumentar el rendimiento y reducir costos.
- Asegurar la retención de clientes y reducir el abandono.
- Asegurar la apertura y acceso a nuevos mercados.

Al igual que las herramientas de minería de datos existen muchas tecnologías CRM bajo licenciamiento GNU-GPL que la PYME puede utilizar para emprender estrategias de marketing a sus clientes. Sin embargo en esta ocasión solo daremos a conocer una de ellas la cual consideramos representa mayores

⁴ Fundamentos para establecer una estrategia CRM. Hideki Erigh Hashimura.

⁵ <http://www.sugarcrm.com/sites/default/files/whitepapers/Fundamentos-para-establecer-una-estrategia-CRM.pdf> (Fecha de consulta: 17/07/2015).

portunidades para el marketing relacional, dicha tecnología es SugarCRM.



SugarCRM es un sistema para la administración con los clientes basado en LAMP (Linux, Apache, MySQL, PHP). Es una aplicación muy completa para negocios de distintos tamaños, está diseñada para facilitar la gestión de ventas, oportunidades, contactos de negocios entre otros. Está disponible en cinco ediciones:

- Sugar Community Edition.
- Sugar Profesional.
- Sugar Corporate.
- Sugar Enterprise.
- Sugar Ultimate.

Ofrece además diferentes opciones de implementación:

- Software como servicio (Saas).
- Windows Azure Services Platform.
- Rackspace Cloud.
- IBM SmartCloud Enterprise.

La implementación de SugarCRM permite la centralización de información y:

- Ofrece a los gerentes visibilidad instantánea de las actividades y oportunidades en curso de equipos.
- Alerta a los gerentes sobre fechas límites cercanas.
- Garantiza que se respete el mismo proceso de ventas en todos los países.
- Estandariza la gestión de informes para permitir la comparación de métricas comunes de rendimiento independientemente de la ubicación geográfica.

SugarCRM es gestión eficaz de la información, la velocidad a la que la información, las personas y los proyectos cambian, requieren de nuevas metodologías para la gestión.

Es una potente herramienta de comunicación en un mundo en el que cada cliente es único, la segmentación y la comunicación personalizada son claves del éxito. Mercados dominados por el rápido consumo, los clientes demandan cada vez más atención y dedicación de recursos.

Tres de las consideraciones fundamentales que se deben tener a la hora de afrontar una implantación CRM son: análisis, metodología y formación constante.

VI. CASOS DE APLICACIÓN

6.1.1 Análisis de clientes potenciales

Existe la necesidad de identificar, entender y mejorar los flujos de trabajo implícitos en los procesos corporativos de ventas que permita facilitar la toma de decisión en la promoción de nuevos productos. En este caso, un plan inefectivo de

producción podría generar como consecuencia la acumulación de stock y la pérdida de utilidades.

La empresa “X” es actualmente una floreciente comercializadora de productos lácteos con un aproximado de 500 clientes, ellos van a introducir al mercado el nuevo “Queso Dip” y desean saber quiénes son clientes potenciales para dicho producto. Para poder recuperar la inversión inicial la empresa se ha focalizado en los clientes excelentes por lo que será necesario identificarlos de la base de datos.

Estaciones para encontrar los clientes potenciales:

- Clientes correctos.
- Relaciones correctas.
- Retención correcta.

Cientes correctos: nos interesa analizar los clientes valiosos que puedan aplicar a nuestra propuesta de negocio, que a menudo se identifican como los clientes que tienen consumo regular en un mes a lo largo del año. Es importante determinar en nuestro análisis identificar los clientes que son leales y rentables.

Relaciones correctas: incluso con una buena elección de clientes es importante desarrollar la relación a largo plazo, el hecho de no conseguirlo puede significar un impacto negativo en lo que esperamos como valor de ganancia. Para cualquier empresa, la relación correcta es aquella que maximiza el valor de vida del cliente.

Retención correcta: la retención efectiva significa retener al cliente correcto, y no así todos los clientes. Se debe prestar atención en las acciones de retención en clientes que poseen un alto valor de vida. Puede ser contraproducente si gastamos demasiados recursos valiosos en el intento de retener a clientes que no son rentables, la estrategia por tanto es enfocarse en quien es realmente rentable para el negocio.

6.1.2 Formulación del modelo

Una vez el problema a ser analizado es adecuadamente identificado, los esfuerzos deben estar direccionados a definir un modelo matemático que represente nuestro sistema. Son varios los factores que afectan e inciden en la elección del modelo, tales como una línea de tiempo, variables de decisión, criterio de evaluación, parámetros numéricos y relaciones matemáticas.

Variables de decisión para el problema identificado:

- Edad
- Sexo
- Salario promedio (Estimado)
- Profesión
- Dependencia de familiares
- Ubicación
- Producto
- Cantidad

- Tamaño de envase
- Frecuencia de compra
- Momentos de compra
- Monto de la compra
- Donde realiza la compra
- Numero de reclamos
- Nivel de satisfacción
- Antigüedad

6.1.3 Desarrollo

Recopilación de datos e integración: una vez que se ha analizado la meta de investigación, la recopilación de datos e información pueden residir en diferentes fuentes y es labor del analista integrarlos para cumplir con el objetivo trazado.

El trabajo de coleccionar la información puede ser arduo sino se cuenta con la tecnología apropiada, mantener la integridad de los datos es un reto para el desarrollo de un modelo de minería de datos. Un data mart independiente puede ayudarnos a simplificar esta tarea, los datos que se utilizan para construirlo provienen de los sistemas operacionales y en algunos casos de fuentes externas. Para el desarrollo de la aplicación del caso en estudio se utilizó una base experimental de 500 clientes bajo formato hoja de cálculo.

Pre procesamiento de datos: el pre procesamiento de datos es una tarea tediosa y obligatoria en el análisis de información, se utiliza principalmente para la fabricación apropiada de análisis mineros, se construye mediante la supresión de registros duplicados y el suministro de datos faltantes para facilitar un análisis de minería. Los principales beneficios del pre procesamiento de datos son la reducción de memoria en las herramientas tecnológicas y la normalización de valores que se almacenan en las bases de datos.

Análisis exploratorio: en esta fase del proceso de minería es donde preparamos la limpieza de datos seleccionando las variables de interés para nuestro análisis. La figura siguiente muestra el resultado de las variables seleccionadas para el desarrollo del caso de aplicación clientes potenciales.

ID	NOMBRE	SEXO	EDAD	ESTADUAL	USUARIO	ESPO	OFICIO	CIUDAD	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
1	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
2	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
3	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
4	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
5	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
6	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
7	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
8	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
9	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR
10	OSCAR	M	23	BOGOTÁ	USUARIO	N	OFICINISTA	BOGOTÁ	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR	INDICADOR

Figura 7: Análisis exploratorio de las variables.

Selección de atributo: los atributos son las propiedades o características de los objetos, por ejemplo el color de ojo de una persona, la temperatura, etc. Los atributos también se conocen como variables. Un conjunto de atributos describe un objeto o

un registro. Los atributos son importantes para la selección de una técnica de minería.

Más adelante en el modelo se describe cómo es que se obtienen los atributos, que son la base para el análisis minero.

Aplicación de la técnica: para el desarrollo del caso de aplicación nos auxiliamos con la herramienta Rapid Miner, como antes se ha descrito en éste documento, Rapid Miner nos permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico.

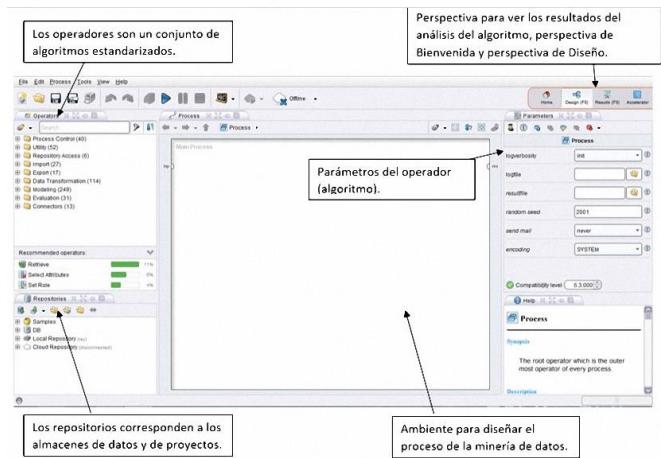


Figura 8: Entorno gráfico de Rapid Miner.

El primer paso a realizar es la carga del archivo en formato hoja de cálculo ya generado en el análisis exploratorio, esto se realiza desde el menú principal – File – Import Data – Import CSV File, tal como lo muestra la siguiente figura.

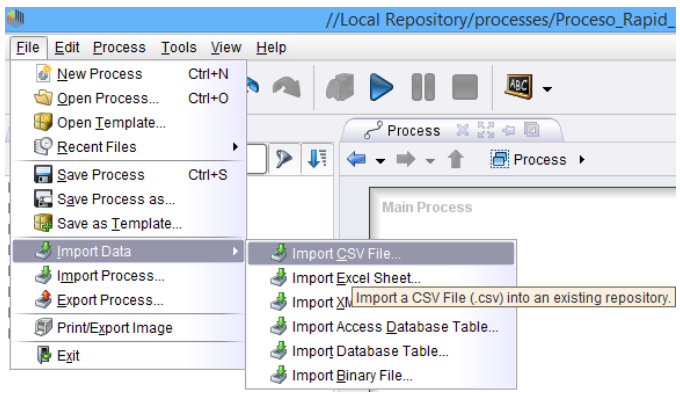


Figura 9: Importar datos desde Rapid Miner.

Un asistente nos guiará para finalizar el procedimiento, es importante mencionar que es aquí donde se da el proceso de selección de atributos. Sin embargo para el análisis final se deben de utilizar únicamente atributos numéricos, por lo que haremos uso de un operador especial en el manejo de atributos que se explica más adelante.

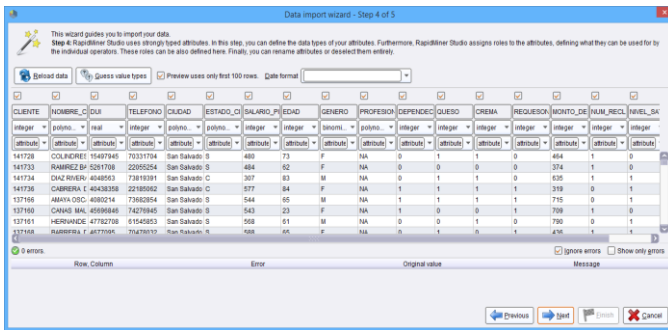


Figura 10: Selección de atributos en Rapid Miner.

Al finalizar el asistente debemos tener el operador inicial como se muestra en la siguiente figura.

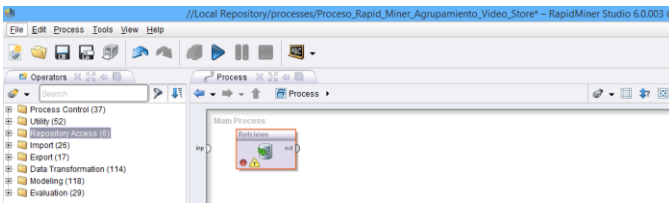


Figura 11: Adición de operador Retrieve en Rapid Miner.

El siguiente paso es agregar el operador Set Role, sirve para modificar el rol de uno o más atributos para su posterior análisis. Se utilizará para elegir las variables definitivas para la demostración de la técnica de minería de datos K-Means.

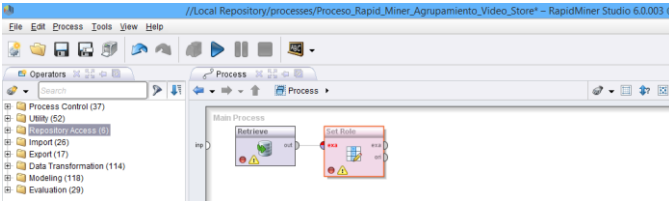


Figura 12: Adición de operador Set Role en Rapid Miner.

La siguiente imagen describe la ventana de edición del operador, el rol definido para todos los atributos seleccionados es "Regular". La interconexión de los operadores debe ser realizada tal como se muestra en las imágenes.

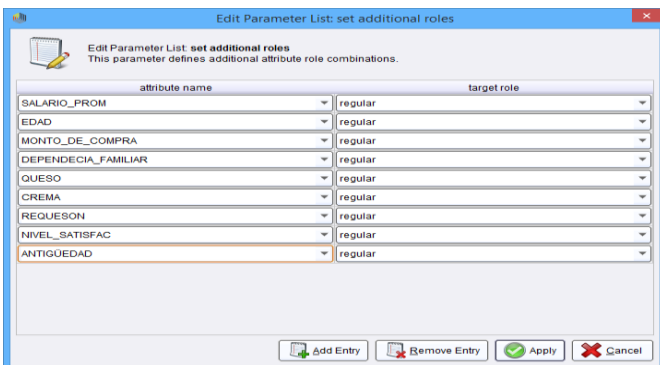


Figura 13: Ventana de edición operador Set Additional Roles en Rapid Miner.

El operador Select Attributes nos permite seleccionar los atributos definitivos para someterlos al set de ejemplo.

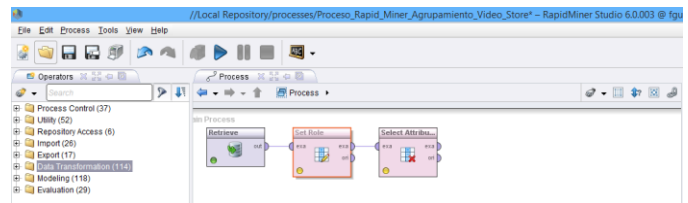


Figura 14: Adición de operador Select Attribute en Rapid Miner.

La ventana de edición del operador se muestra a continuación.

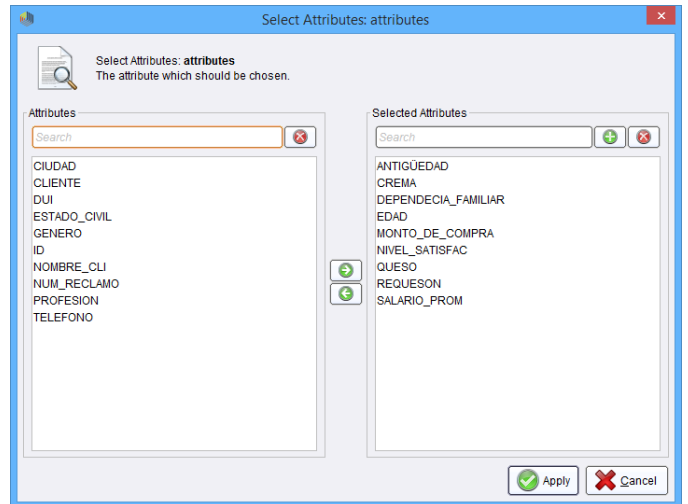


Figura 15: Ventana de edición de operador Select Attribute en Rapid Miner.

Una vez definidos los atributos es necesario utilizar el operador Normalize. La técnica de minería de datos K Means utiliza la distancia euclidiana como métrica, por tanto todos los atributos deben poseer una misma escala para poder realizar análisis muy aproximados de la información. En otras palabras podemos utilizar este operador cuando tengamos información con escalas desiguales asegurando que las distancias calculadas para generar los grupos no sean influenciadas por factor escala.

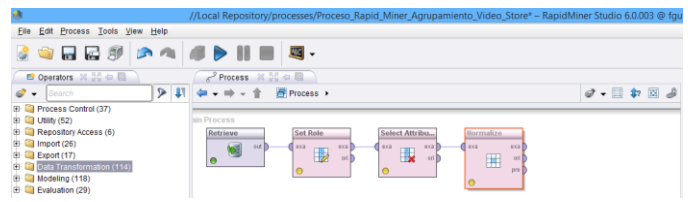


Figura 16: Adición de operador Normalize en Rapid Miner.

El operador clustering es el que se encarga de ejecutar la técnica propiamente. K Means es una técnica de minería de

agrupamiento muy conocida, sigue una forma fácil para dividir una base de datos dada en K grupos⁶.

La idea principal con la técnica es definir K centroides (uno para cada grupo generado) y luego tomar cada punto de la base de datos y situarlo en la clase de su centroide más cercano y así sucesivamente se recalculan los centroides de cada grupo para redistribuir los objetos según el centroide resultante más cercano. El proceso se repite hasta que ya no hay cambio en los grupos. La figura siguiente muestra la interconexión del operador.

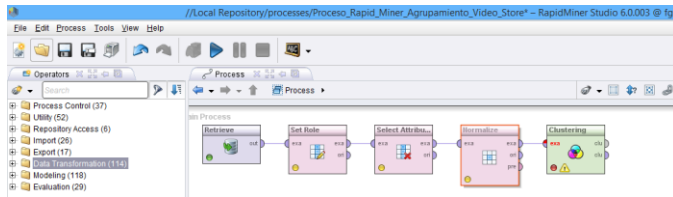


Figura 17: Adición de operador Clustering en Rapid Miner.

En nuestro ejemplo vamos a generar dos agrupaciones, parametrizando la variable k como se muestra a continuación:

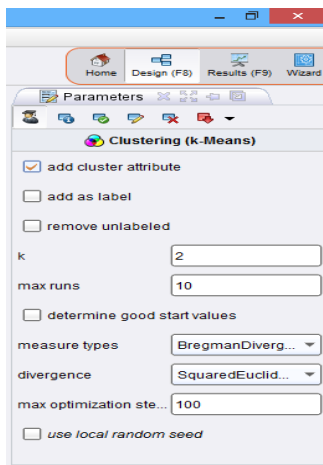


Figura 18: Parametrización de la variable K en Rapid Miner.

El add cluster attribute es seleccionado para poder exportar la data final a una hoja de cálculo, mientras que max runs define las repeticiones para el recalcado de centroides.

Finalmente el operador Cluster Distance Performance toma la información de los centroides calculados, los utiliza como entrada de información y evalúa el desempeño del modelo basado en los métodos de agrupación. Con la conexión de éste operador se concluye la construcción del modelo para

segmentar clientes con el objetivo de identificar aquellos más valiosos para nuestras campañas de marketing.

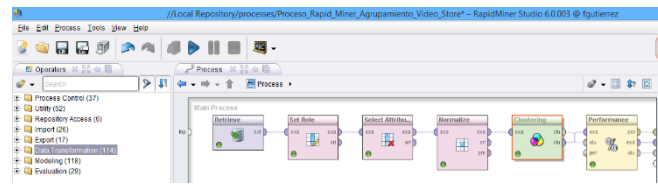


Figura 19: Modelo resultante en Rapid Miner.

6.1.4 Interpretación de resultados

Una de las ventajas que ofrece Rapid Miner es que provee varias opciones para analizar los datos. La descripción de la vista de resultado nos muestra el agrupamiento resultante con dos grupos diferenciados (K) que han sido especificados en el desarrollo del modelo. La figura siguiente describe el número de elementos pertenecientes a cada grupo.

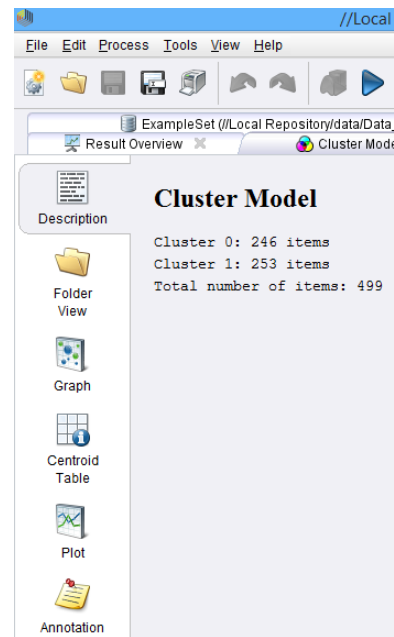


Figura 20: Resultado de agrupamiento en modo text view en Rapid Miner.

La vista de carpeta nos permite una mejor perspectiva sobre los integrantes de cada grupo resultante. En la figura 21 se detallan los ID de cada cliente por grupo.

⁶ Sánchez, Pascual. "Algoritmos de agrupamiento" (2007). Departamento de lenguajes y sistemas informáticos. Web 20 de agosto de 2015.

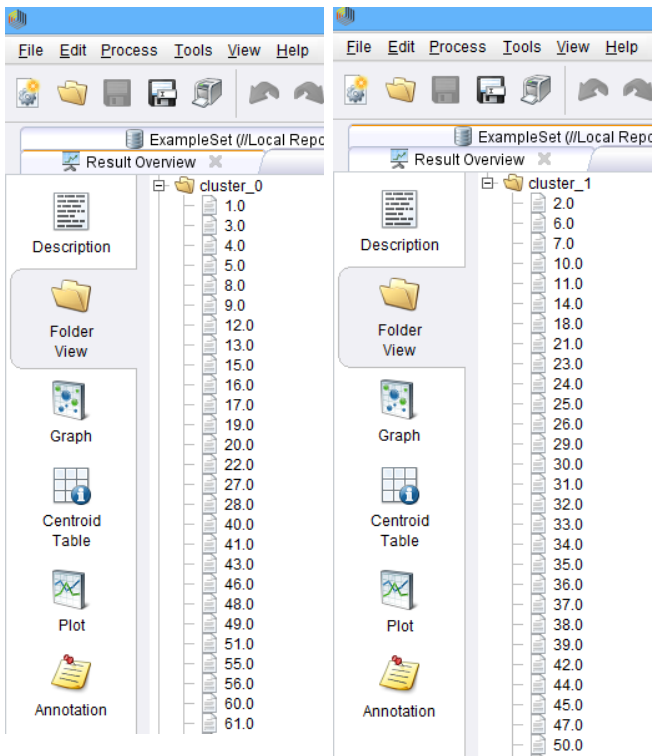


Figura 21: Resultado de agrupamiento por ID cliente en Rapid Miner.

La tabla de centroides se muestra en la figura 22, representa en sí los vectores de medias de todas las variables utilizadas en el agrupamiento. El centroide resume las características de ambos grupos.

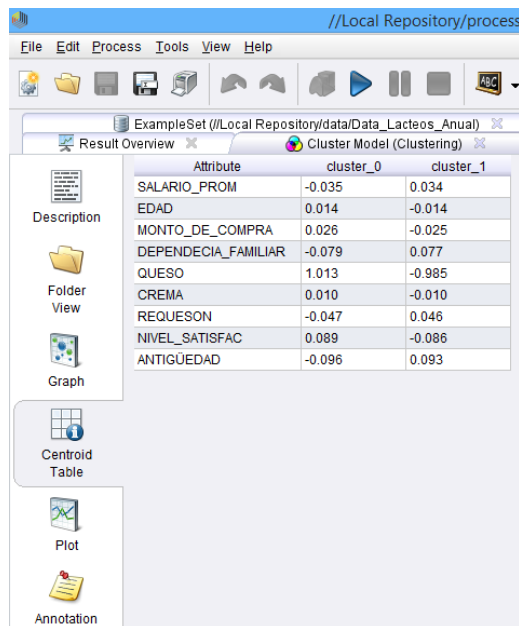


Figura 22: Resultado tabla de centroides en Rapid Miner.

La vista de Plot (Figura 23) nos muestra un gráfico de coordenadas paralelas, la elección del grupo objetivo para la campaña de marketing, se define analizando estos resultados. Para el caso de aplicación que llevamos en curso el grupo 0 ofrece mejores posibilidades de clientes con mejor opción para la aceptación del nuevo producto.

Aunque el grupo no precisamente cuenta con los mejores salarios, pero si muestra alta tendencia en la compra de queso, por lo que la opción de queso DIP podría ser ofrecida con alta expectativa de aceptación.

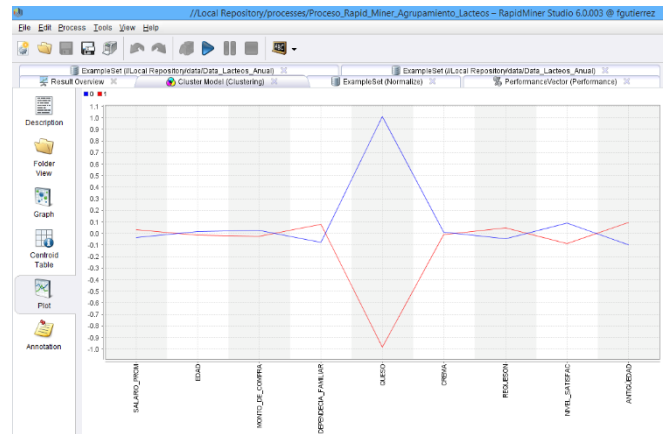


Figura 23: Resultado en modo centroid plot view en Rapid Miner.

6.2.1 Análisis de abandono de clientes

Un aspecto muy importante que toda organización debe tener en consideración es la tasa de abandono de los clientes, si los clientes no están satisfechos cambian y en muchas ocasiones no protestan convirtiéndose así en una problemática ya que se desconocen las causas.

Rank Xerox & Carlson Marketing Group

“Un incremento del 5% en la fidelidad de los clientes puede producir un aumento del 25% al 85% de los beneficios”

Harvard Business School

“Un incremento del 5% en la fuga de clientes puede producir una reducción del 30% en los beneficios”

A continuación se plantea un análisis de casos de abandono de clientes analizado con la herramienta WEKA a través del método de clasificación e implementando la técnica de árboles de decisión.

Los datos han sido extraídos del estudio de *“Segmentación de clientes del comercio minorista en función de los beneficios de la relación minorista-consumidor”*⁷.

⁷ Berenguer, Gloria; Gil Saura, Irene; Ruiz Molina, María “Segmentación de clientes del comercio minorista en función de los beneficios de la relación minorista-consumidor” (2007).

FICHA TÉCNICA	
Establecimientos:	Alimentación Calzado Electrodomésticos Muebles
Tamaño muestral:	100 consumidores
Diseño muestral:	Datos transaccionales
Elaboración:	2015
Idoneidad de la muestra:	Análisis de reclamos y abandono
Análisis previo:	Análisis de antigüedad Análisis de reclamos
Técnicas estadísticas:	Análisis de conglomerados Análisis predictivo
Programa estadístico:	WEKA 3.7.12

Cuadro 1: Ficha técnica de datos analizados.

DATOS GENERALES DE LA MUESTRA		
Establecimientos:		Registros
Alimentación	-----	25
Calzado	-----	22
Electrodomésticos	-----	27
Muebles	-----	26
Total	-----	100
Género		
Hombre	-----	47
Mujer	-----	53
Total	-----	100
Edad		
18-25	-----	6
26-35	-----	16
36-45	-----	31
46-55	-----	13
56-65	-----	34
Total	-----	100
Nivel de estudio		
Sin estudio	-----	18
Primarios	-----	27
Secundarios	-----	40
Universitarios	-----	15
Total	-----	100
Antigüedad de la relación del cliente con la tienda		
0-1	-----	20
2-5	-----	28
6-10	-----	24
11-15	-----	28
Total	-----	100

Cuadro 2: Ficha técnica de datos analizados.

A estos datos de la muestra se suman las variables dependientes de “reclamo” y “abandono” para realizar el análisis quedando de la siguiente manera:

Variables de análisis para el problema identificado:

- Establecimiento
- Género
- Edad
- Nivel de estudio
- Antigüedad
- Reclamos
- Abandono

ESTABLECIMIENTO	GÉNERO	EDAD	NIVEL DE ESTUDIO	ANTIGÜEDAD	RECLAMOS	ABANDONO
Alimentación	Hombre	18-25	Sin estudio	0-1	1	no
Electrodomésticos	Mujer	56-65	Primarios	2-5	0	si
Muebles	Mujer	36-45	Secundarios	6-10	2	no
Calzado	Hombre	46-55	Universitarios	11-15	3	no
Alimentación	Mujer	56-65	Primarios	6-10	1	si
Alimentación	Hombre	46-55	Primarios	0-1	0	no
Muebles	Mujer	36-45	Secundarios	0-1	3	si
Calzado	Hombre	18-25	Sin estudio	6-10	2	no
Alimentación	Mujer	46-55	Secundarios	0-1	1	no
Electrodomésticos	Mujer	56-65	Primarios	2-5	1	si
Muebles	Hombre	26-35	Secundarios	11-15	3	no
Calzado	Mujer	56-65	Sin estudio	6-10	0	si
Alimentación	Hombre	18-25	Primarios	11-15	0	si
Alimentación	Mujer	36-45	Sin estudio	0-1	1	no
Electrodomésticos	Mujer	18-25	Sin estudio	2-5	0	si
Alimentación	Mujer	56-65	Sin estudio	11-15	1	si
Muebles	Hombre	36-45	Secundarios	6-10	1	no
Electrodomésticos	Hombre	18-25	Sin estudio	11-15	1	si
Electrodomésticos	Mujer	56-65	Secundarios	6-10	0	no
Calzado	Hombre	26-35	Sin estudio	2-5	1	no

Figura 24: Análisis exploratorio de las variables para determinar el abandono.

Una vez se tienen las variables definidas y los datos recopilados por medio de las transacciones, se procede hacer uso de la herramienta para el análisis.

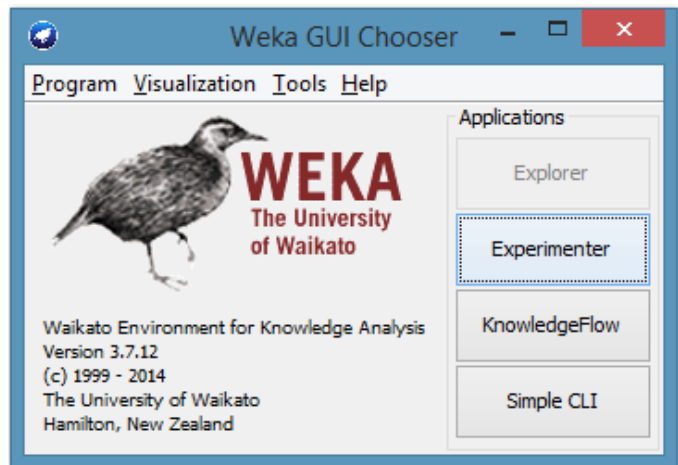


Figura 25: Interfaz principal de WEKA 3.7.12.

La herramienta contiene cuatro opciones para la explotación de los datos, para nuestro ejemplo utilizaremos la primera opción a modo exploratorio para la lectura de los registros seleccionados anteriormente.

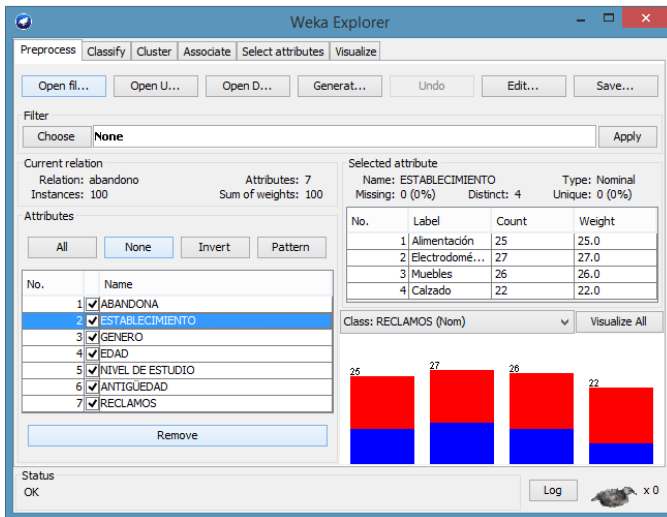


Figura 26: Selección de datos en archivo csv.

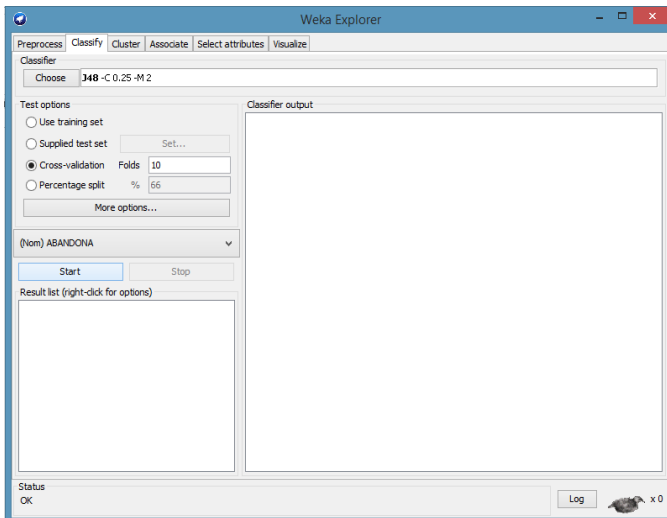


Figura 27: Selección del método y técnica de minería de datos.

Es aquí donde seleccionamos el método de clasificación y además la técnica algorítmica para el análisis de las variables que como se mencionó al inicio se utilizará árboles de decisión, para lo cual WEKA dispone de la notación J48-C0.25-M2 para la aplicación de ésta técnica.

Además se debe seleccionar la variable dependiente sobre la cuál determinará la clasificación por medio del procesamiento y de manera híbrida la construcción del árbol, siendo un aprendizaje supervisado estas técnicas permitirán la predicción del dato de abandono.

=== Run information ===

```

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    abandono
Instances:   100
Attributes:  7
             ABANDONA
             ESTABLECIMIENTO
             GENERO
             EDAD
             NIVEL DE ESTUDIO
             ANTIGÜEDAD
             RECLAMOS
Test mode:   10-fold cross-validation

```

=== Classifier model (full training set) ===

=== Classifier model (full training set) ===

J48 pruned tree

```

-----
NIVEL DE ESTUDIO = Sin estudio
|  EDAD <= 46: no (12.0/4.0)
|  EDAD > 46: yes (6.0/1.0)
NIVEL DE ESTUDIO = Primarios: yes (27.0/10.0)
NIVEL DE ESTUDIO = Secundarios
|  ESTABLECIMIENTO = Alimentación
|  |  EDAD <= 46: no (5.0)
|  |  EDAD > 46: yes (3.0/1.0)
|  ESTABLECIMIENTO = Electrodomésticos
|  |  ANTIGÜEDAD <= 6: yes (5.0/1.0)
|  |  ANTIGÜEDAD > 6: no (2.0)
|  ESTABLECIMIENTO = Muebles
|  |  ANTIGÜEDAD <= 0: yes (2.0)
|  |  ANTIGÜEDAD > 0: no (15.0/2.0)
|  ESTABLECIMIENTO = Calzado
|  |  RECLAMOS = true: no (3.0/1.0)
|  |  RECLAMOS = false
|  |  |  ANTIGÜEDAD <= 1: no (2.0)
|  |  |  ANTIGÜEDAD > 1: yes (3.0)
NIVEL DE ESTUDIO = Universitarios
|  RECLAMOS = true: no (4.0/1.0)
|  RECLAMOS = false
|  |  ANTIGÜEDAD <= 0: no (2.0)
|  |  ANTIGÜEDAD > 0
|  |  |  ANTIGÜEDAD <= 1: yes (2.0)
|  |  |  ANTIGÜEDAD > 1
|  |  |  |  EDAD <= 36: yes (4.0/1.0)
|  |  |  |  EDAD > 36: no (3.0)

```

Figura 28: Resultado del análisis WEKA.

Dicho resultado se lee de la siguiente manera a través del árbol de decisión.

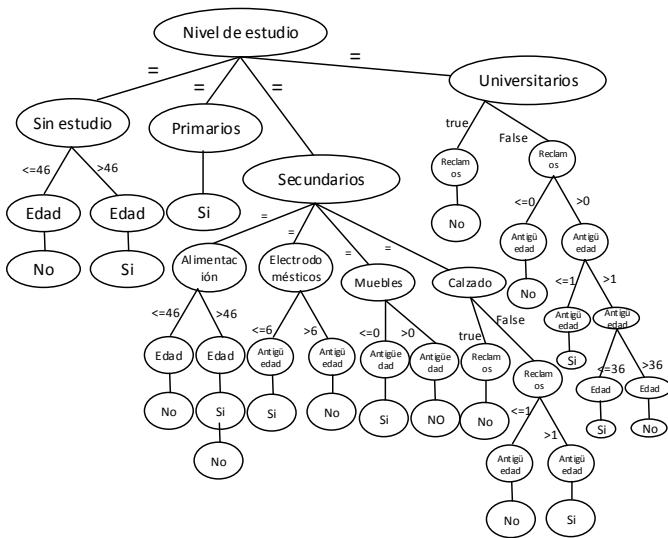


Figura 29: Árbol de decisión del análisis WEKA.

De manera muy similar al primer caso de aplicación se han utilizado los mismos componentes de agrupación de datos a diferencia que no se necesitó el componente de “Select Attributes”.

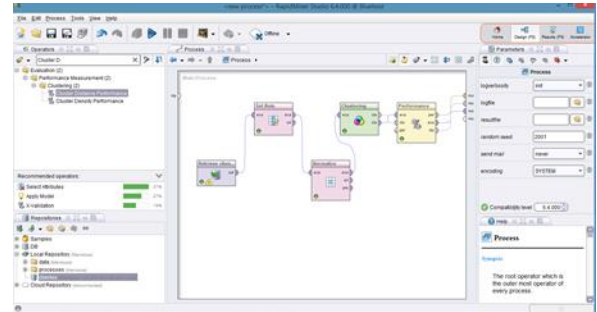


Figura 31: Componentes de agrupamiento de datos.

6.2.2 Interpretación de resultados

Para la determinación del análisis es necesario tener en claro que la variable abandono es la dependiente de aquellas que consideramos influyen o determinan que ésta se cumpla o pronostique. Se puede apreciar que existen personas que a pesar que hayan hecho reclamos no abandonarán la empresa, pero que deben cumplir con cierta edad o antigüedad de ser clientes. Entonces el provecho de este análisis es conocer las características de los individuos y entre qué rangos que plantean el problema desde distintas perspectivas, analizando todas las posibles soluciones ayudando a tomar las mejores decisiones por medio de las probabilidades existentes.

6.3.1 Una estrategia efectiva para retener los clientes

Para la retención de clientes se requiere tener muy en cuenta qué perfiles de individuos se necesita retener, retomando los datos anteriores para el árbol de decisión se crearán clúster para conocer qué establecimiento tiene mayor monto de compras, cantidad de usuarios, antigüedad y las edades promedio de los clientes. Esto permitirá como se planteó en la retención de clientes enfocar nuestros esfuerzos aquellos clientes que generan mayores réditos y saber quiénes son para poder fidelizarlos.

ESTABLECIMIENTO	EDAD	ANTIGÜEDAD	RECLAMOS	MONTO CONSUMO
Alimentación	18	1	1	971
Electrodomésticos	65	2	0	271
Muebles	36	6	2	614
Calzado	46	2	3	457
Alimentación	56	10	1	4567
Alimentación	55	0	0	898
Muebles	36	0	3	310
Calzado	25	6	2	884
Alimentación	55	1	1	979
Electrodomésticos	56	2	1	189
Muebles	35	11	3	124
Calzado	56	6	0	369
Alimentación	18	11	0	1247
Alimentación	36	1	1	245
Electrodomésticos	18	2	0	245
Alimentación	56	5	1	2457
Muebles	45	6	1	4578
Electrodomésticos	25	15	1	979
Electrodomésticos	56	10		475

Figura 30: Análisis de consumo por establecimiento.

Los resultados permiten conocer los tres grupos y las respectivas cantidades de ítems que los conforman.

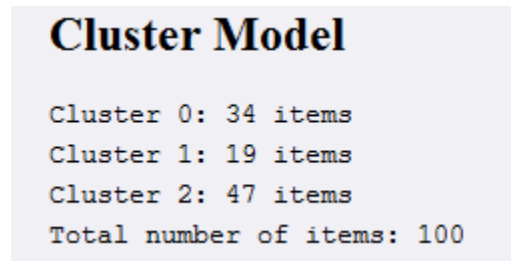


Figura 32: Cantidad de ítems por clúster.

La gráfica siguiente que es resultado del agrupamiento, nos permite conocer qué a pesar que el Clúster 1, tiene menor cantidad de ítems, es aquel que mayor consumo genera.

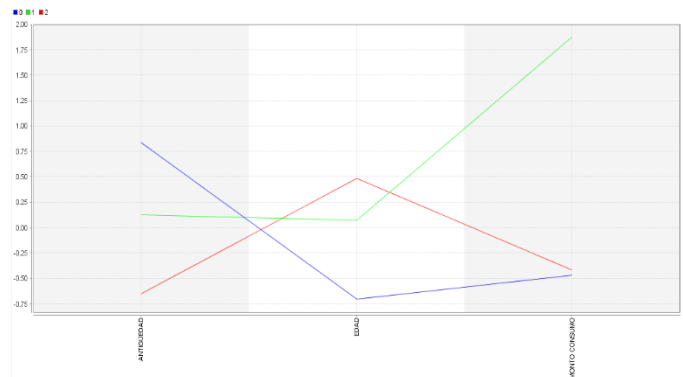


Figura 33: Cantidad de ítems por clúster.

Esto de igual manera permitirá redireccionar esfuerzos a los establecimientos en los cuales se obtienen mayores ventas y comprobar además ¿Sí a altos montos se tienen mayores ganancias? Estas operaciones se podrán realizar con un análisis de costos beneficios, sin embargo la finalidad del caso es crear las agrupaciones y mostrar al lector el cómo puede conocer sus segmentos de mercados en edades, antigüedad y consumo por

establecimientos de ventas. Para conocer los establecimientos bastará con filtrar los datos por cluster como se muestra en la siguiente figura.

ESTABLECIMIENTO	RECLAMOS	cluster	ANTIGÜEDAD	EDAD	MONTO CONSUMO
Alimentación	1	cluster_0	-0.818	-2.014	-0.522
Electrodomésticos	0	cluster_2	-0.579	1.547	-0.766
Muebles	2	cluster_0	0.378	-0.650	-0.647
Calzado	3	cluster_2	-0.579	0.108	-0.701
Alimentación	1	cluster_1	1.334	0.865	0.729
Alimentación	0	cluster_2	-1.057	0.790	-0.548
Muebles	3	cluster_2	-1.057	-0.650	-0.752
Calzado	2	cluster_0	0.378	-1.484	-0.553
Alimentación	1	cluster_2	-0.818	0.790	-0.520
Electrodomésticos	1	cluster_2	-0.579	0.865	-0.794
Muebles	3	cluster_0	1.573	-0.726	-0.817
Calzado	0	cluster_2	0.378	0.865	-0.732
Alimentación	0	cluster_0	1.573	-2.014	-0.426
Alimentación	1	cluster_2	-0.818	-0.650	-0.775
Electrodomésticos	0	cluster_0	-0.579	-2.014	-0.775

Figura 34: Resultado de la creación de clúster.

VII. CONCLUSIONES

Las infinitas formas de conquista y posicionamiento de mercado, involucran estrategias, métodos, técnicas, herramientas y tecnologías cada vez más evolucionadas, que demandan de las empresas establecer sistemas de información que permita una penetración de mercado. Por tal razón consideramos de mucha importancia el conocimiento y aplicación de la minería de datos, en todas las empresas que dedican sus esfuerzos a la comercialización de bienes y servicios.

Esta propuesta representa un desafío en el buen término que la PyME debe asumir, no por simple hecho de adquirir tecnología y aumentar sus costes, sino el cambio de paradigma de un marketing operativo a uno relacional. Las actividades comerciales son cada vez más exigentes, no solo con los clientes, sino además con nuestros proveedores en el establecimiento de canales de comunicación y distribución a los cuales en muchas ocasiones debemos ajustarnos con nuestros procesos.

Se han argumentado los resultados que tienen las implicancias de estas prácticas mineras, en las gestiones comerciales y con un valor agregado a la PyME ya que se utilizaron tecnologías Open Source. Esto con la finalidad de animar a este sector para beneficiarse en su implementación para las actividades del marketing relacional.

Sabemos que cualquiera que sea el problema a resolver, no existe una única técnica para solucionarlo, sino que puede ser abordado a través de distintas aproximaciones, a medida la PyME se interese y utilice estas valiosas herramientas podrá

sacar el mejor provecho para sus clientes y la organización. Creando relaciones duraderas y beneficiosas para ambas partes.

REFERENCIAS

- [1] Pérez López, César; Santín González, Daniel.
MINERÍA DE DATOS TÉCNICAS Y HERRAMIENTAS
Editorial THOMSON Paraninfo, S.A., Madrid España 2008, 2da Edición.
- [2] Nettleton David F.
TÉCNICAS PARA EL ANÁLISIS DE DATOS CLÍNICOS
Editorial Díaz de Santos, España 2005, 1ra Edición.
- [3] Kotler, Philip; Keller Lane, Kevin.
DIRECCIÓN DE MARKETING
Editorial PEARSON EDUCATION, México 2006, Duodécima edición.
- [4] Jane Smith.
RETENER Y FIDELIZAR A LOS CLIENTES
Editorial GESTION 2000, España 2005, 1ra Edición.
- [5] Guitman, Lawrence J.
FUNDAMENTOS DE ADMINISTRACIÓN FINANCIERA
Editorial HARLA S.A., México 1992, 3ra Edición.
- [6] Guitman, Joseph P.; Gordon W., Paul.
ADMINISTRACIÓN DE MERCADEO. ESTRATEGIAS Y PROGRAMAS.
Editorial McGRAW-HILL, México 1984, 1ra Edición.
- [7] Baca Urbina, Gabriel.
EVALUACIÓN DE PROYECTOS
Editorial McGRAW-HILL, México 2001, 4ta Edición.
- [8] Liberos, Eduardo; Somalo, Ignacio; Gil, Juan; García del Poyo, Rafael; Merino, Juan Antonio.
LIBRO DEL COMERCIO ELECTRÓNICO.
Editorial ESIC, Madrid España 2011, 2da Edición.
- [9] Armario, Enrique Martín.
MARKETING RELACIONAL
Editorial ESIC, Madrid España 1999, 1ra Edición.
- [10] Martin, Christopher; Payne, Adrian; Ballantyne David.
MARKETING RELACIONAL, Integrando la calidad, el servicio al cliente y el marketing.
Editorial DÍAZ DE SANTO, S.A. Madrid España 1994, 1ra Edición.
- [11] Lucio Mera, Elena.
TÉCNICAS DE MARKETING. IDENTIFICAR, CONQUISTAR Y FIDELIZAR CLIENTES.
Editorial IDEAS PROPIAS, España 2005, 1ra Edición.
- [12] Escudero Serrano, M^a José.
GESTIÓN COMERCIAL Y SERVICIO DE ATENCIÓN AL CLIENTE.

Editorial PARANINFO, S.A., Madrid España 2011, 2da Edición.

[13] Vieira Braga, Luis Paulo; Ortiz Valencia, Luis Iván; Ramírez Carvajal, Santiago Segundo.
INTRODUCCIÓN A LA MINERÍA DE DATOS
Editorial E-PAPERS, Brasil Río de Janeiro 2010, 1ra Edición.

[14] Woodburn, Diana, McDonald, Malcolm.
EL MANAGEMENT DE LAS CUENTAS CLAVE –
Como seleccionar, tratar y retener a los clientes más rentables.
Editorial GRANICA, Buenos Aires 2013, 1ra Edición.

[15] Castelló Martínez, Araceli
ESTRATEGIAS EMPRESARIALES EN LA WEB 2.0 -
las redes sociales online.
Editorial CLUB UNIVERSITARIO, España 2010, 1ra Edición.