



Detección de Patrones Transaccionales en Cuentahabientes Bancarios: una Aplicación de la Metodología SEMMA

Acerca de los autores...

¹ Alumna de la Maestría en Ingeniería en Sistemas Computacionales, Tecnológico de Estudios Superiores de Ecatepec.
^{a, b y c} Docente de la División de Ingeniería en Sistemas Computacionales, Tecnológico de Estudios Superiores de Ecatepec.
lamatinezueu@hotmail.com, acerterus_paribus@hotmail.com, becorona@tese.edu.mx, cajja_mx@yahoo.com

Alma Leonor Martínez Eugenio¹, Edgar Efrén López Torres^a
Edgar Corona Organiche^b, Abraham Jorge Jiménez^c

Resumen

Este trabajo propone un modelo analítico con base en la Metodología SEMMA y la hibridación de las técnicas de minería de datos de clustering, análisis multivariado y regresión lineal, con el fin de identificar los segmentos de productos bancarios en los que pueden ser clasificados los cuentahabientes. Los resultados obtenidos muestran una clasificación de cuentahabientes bancarios con una eficiencia del 98%.

Palabras Clave: Reconocimiento de patrones, minería de datos, clustering, regresión, metodología SEMMA.

Introducción

En la actualidad, las instituciones bancarias suelen moverse en un entorno que cambia constantemente, es por esto que deben tener la capacidad de adaptarse, aprender, resolver problemas y establecer nuevos métodos para su solución [1]. La minería de datos (MD) explora y extrae información oculta en datos agrupados; en el medio financiero se utiliza como una herramienta para conocer mejor a los clientes y sus hábitos de consumo. Existen empresas que están a la vanguardia gracias al análisis del perfil de su mercado, lo cual les permite segmentar y ofrecer productos que se adapten mejor a las necesidades del mismo. Esto sirve para ubicar puntos de servicio, mejorar fuerzas de venta, identificar consumidores habituales, entre otros aspectos. Lo anterior permite mejorar la toma de decisiones para el mercadeo, la venta de productos y la mejora de servicios basados en la realidad de los mercados [2]

En el presente trabajo se propone un modelo con base en la metodología SEMMA (Sample, Explore, Modify, Model, Assess) y utilizando las técnicas de clustering [3], análisis multivariado [4] y regresión lineal [5] para identificar patrones de transaccionalidad de clientes bancarios con el fin de abstraer hábitos de consumo que facilitará a las instituciones bancarias generar estrategias de venta enfocadas a necesidades reales.

1. Metodología SEMMA

SEMMA (por sus siglas en inglés) es un acrónimo de Sample, Explore, Modify, Model, and Assess que significa: Seleccionar, Explorar, Modificar, Modelar, y Evaluar. Consiste en una serie de pasos secuenciales desarrollados por SAS Institute Inc., uno de los mayores productores de software para estadísticas e inteligencia de negocio; para seleccionar, explorar y modelar grandes cantidades de datos, con el fin de descubrir patrones implícitos en el conjunto de datos analizados. La metodología está orientada a la aplicación de MD para organizar de forma lógica el sistema de herramientas funcionales de los productos de SAS. La ejecución de cada una de sus fases no está descrita de forma rígida, lo que quiere decir que no es necesario terminar una fase antes de comenzar otra (Figura 1). Las fases de la metodología SEMMA se describen a continuación:

Seleccionar: Comienza con el muestreo de datos, seleccionando el conjunto de datos para el modelado. El conjunto de datos debe ser lo suficientemente grande para contener suficiente información para recuperar, pero lo adecuadamente pequeño para ser utilizado de manera eficiente.

Explorar: Corresponde a la comprensión de los datos mediante el descubrimiento de relaciones previstas y no previstas entre las variables, y también anomalías, con la ayuda de la visualización de datos.

Modificar: Contiene métodos para seleccionar, crear y transformar variables en preparación para el modelado de datos.

Modelar: Consiste en la aplicación de diversas técnicas de modelado de MD en las variables preparadas para crear modelos que posiblemente proporcionen el resultado deseado.

Evaluar: Es el resultado del modelado, muestra la confiabilidad y utilidad de los modelos creados.

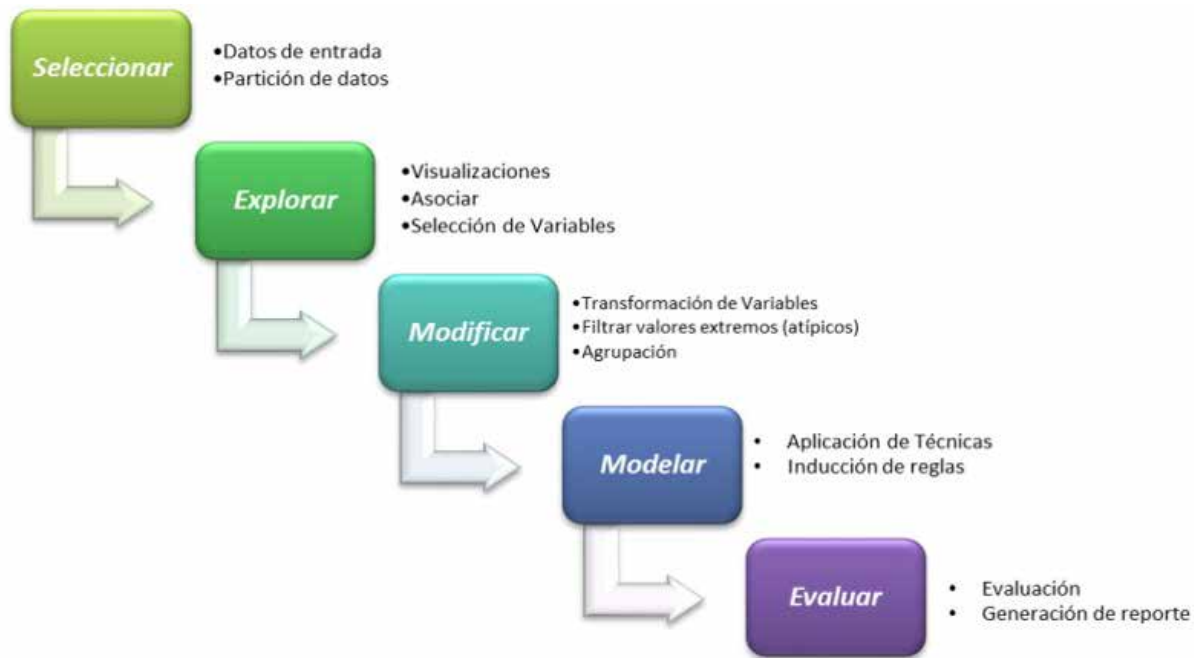


Figura 1

Metodología SEMMA.

2. Modelo Propuesto

La generación de un modelo híbrido surge ante la necesidad de perfilar de forma correcta los cuentahabientes bancarios, ya que se ha identificado que la gran mayoría de ellos cancelan sus productos debido a que las ofertas de los mismos no cumplen con sus necesidades y/o expectativas.

A continuación, se describe el desarrollo del modelo que permite la detección de patrones de transaccionalidad de cuentahabientes bancarios, mediante la hibridación de las técnicas de minería de datos de clustering, análisis multivariado y regresión lineal, para identificar patrones de transaccionalidad de clientes bancarios, conocer sus hábitos de consumo y facilitar a las instituciones bancarias el generar estrategias de venta enfocadas a necesidades reales. La Figura 2 representa las fases que conformarán el modelo analítico.

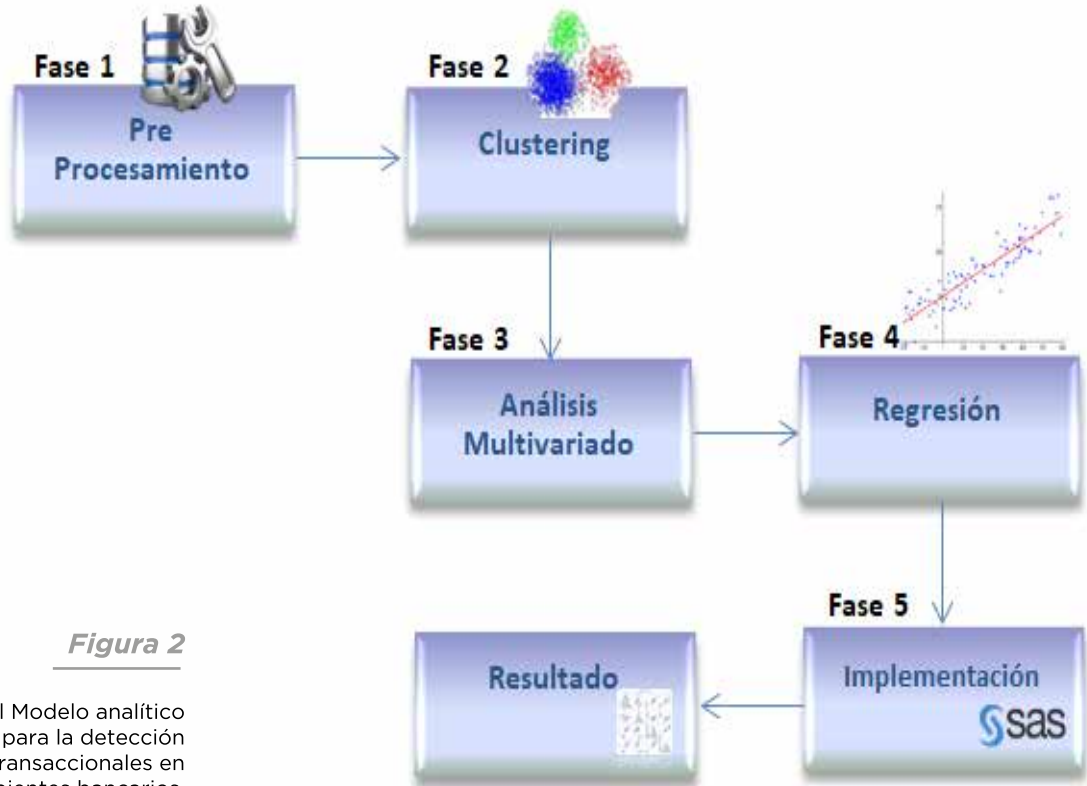


Figura 2

Estructura del Modelo analítico propuesto para la detección de patrones transaccionales en cuentahabientes bancarios.

El modelo toma como universo a los cuentahabientes del producto más rentable de una institución bancaria, para poder medir los principales factores que impulsan el crecimiento del mismo. A continuación, se describe lo que se hizo en cada fase del modelo:

Fase 1. Pre procesamiento. Consiste en la obtención y preparación de los datos relacionados a las cuentas del producto a analizar; engloba las fases de selección y exploración del modelo SEMMA.

En el proceso de **Selección** se extrae de la tabla de Cuenta (Figura 3); en este caso se utilizará el producto más rentable, las cuales deberán ser sólo cuentas activas (Estatus_Cta = Activa dando lugar a la tabla resultado, a la cual llamaremos Cuenta_Mtra. Los productos no deben ser mezclados, ya que las reglas de negocio pueden cambiar de un producto a otro y esto estaría afectando directamente en del clustering.



Figura 3

Selección del Producto.

Una vez obtenida la tabla con las cuentas del producto a analizar, ahora llamada Cuenta_Mtra, es importante identificar los grupos a los que pertenecen los cuentahabientes de acuerdo con su nivel económico; para ello se deben considerar los saldos promedio de las cuentas; los saldos requeridos (Figura 4) son: saldo vista (Sdo_Vista) y saldo integral (Sdo_Int). La figura 4, muestra el proceso de extracción de la cuenta.

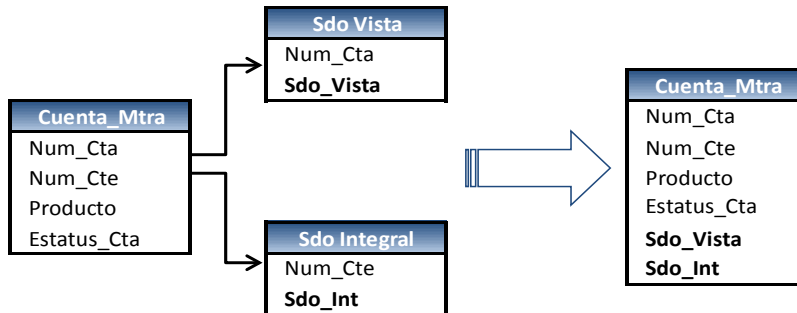


Figura 4

Integración de Saldos.

Sdo_Vista es el saldo promedio en una cuenta del cliente y se obtiene con la suma de los saldos diarios de la cuenta, dividido entre el número de días que conforma el mes (Figura 5).

Cuenta_Mtra																														
Días del Mes																														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo	Sdo
$\text{Sdo_Vista} = \text{Sdo_Promedio_Mes} = (\text{Sdo Dia 1} + \text{Sdo Dia 2} + \text{Sdo Dia 3} + \dots + \text{Sdo Dia 31})/31$																														

Figura 5

Cálculo del Sdo_Vista.

El cálculo del Sdo_Int (Figura 5) se obtiene mediante la suma de los saldos a fin de mes de las inversiones a plazo, la inversión vista, los créditos personales, las hipotecas, las líneas de sobregiro, de préstamo de tarjeta, así como la tarjeta de crédito y banca electrónica que tenga el cliente.

Cliente		
Num Cte	Producto	Sdo Fin Mes
12345	Inversiones a plazo	\$ 0,000.00
12345	Inversión vista	\$ 0,000.00
12345	Créditos personales	\$ 0,000.00
12345	Hipotecas	\$ 0,000.00
12345	Líneas de sobregiro	\$ 0,000.00
12345	Préstamo de tarjeta	\$ 0,000.00
12345	Tarjeta de crédito	\$ 0,000.00
12345	Banca electrónica	\$ 0,000.00
Sdo_Int		Σ \$ 0,000.00

Figura 6

Cálculo del Sdo_Int.

Para poder identificar el perfil de los cuentahabientes bancarios y las variables que influyen en el nivel económico, es importante considerar los datos relacionados al Cliente como son: la edad (Edad), el sexo (Sexo), el estado civil (Edo_Civil), la zona geográfica en la que habita (Edo_dom), el grado de estudio (Niv_Estudios) y ocupación (Ocupacion). Para obtener la información antes mencionada, se requiere relacionar las tablas: Cuenta_Mtra y Cliente (Figura 7).

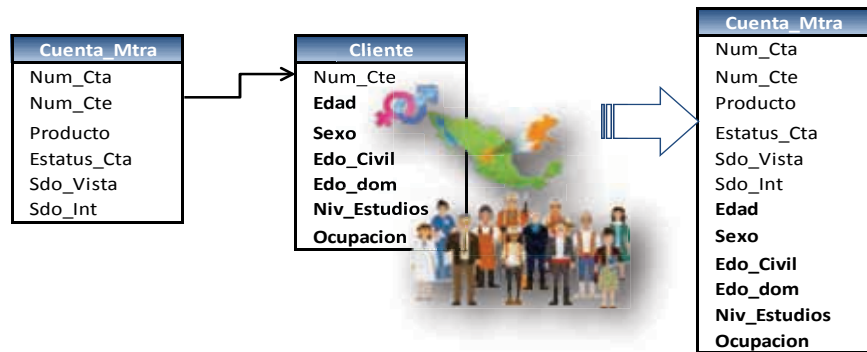


Figura 7

Perfil del Cliente.

De las cuentas incluidas en la base resultante Cuenta_Mtra, debemos clasificar a los cuentahabientes bancarios de acuerdo con el tipo de transacciones que realizan; para ello, se obtendrán las transacciones realizadas identificando: el monto promedio, el número de transacciones, y el tipo de comercio en donde más está transaccionando (Figura 8). La información anterior permitirá identificar las principales actividades que realizan los cuentahabientes.

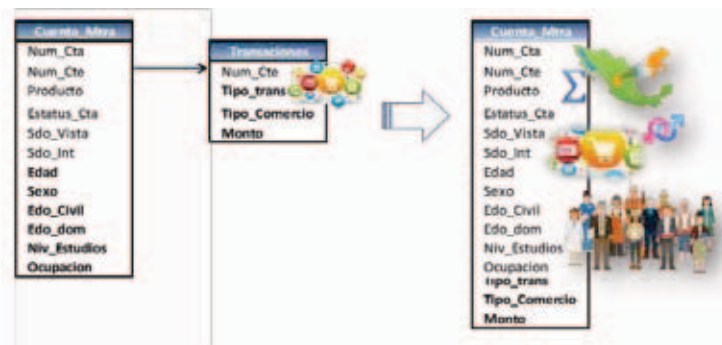


Figura 8

Transaccionalidad del Cliente.

La base que resulte (Figura 9) es la que servirá de insumo para el modelo analítico propuesto



Figura 9

Base de datos que servirá de insumo para el modelo analítico.

Una vez obtenido el insumo principal, se inicia con la fase **Exploración**, en la cual se hace un recorrido a través de los datos extraídos en la muestra para identificar, detectar y depurar datos atípicos (con base en el modelo de caja [6]) -aplicado a Sdo_Vista y Sdo_Int; ayudando a simplificar los procesos de descubrimiento de información en fases siguientes del proceso.

En este punto del proceso, la exploración se puede realizar a través de medios visuales, aunque muchas veces no es suficiente este método, es por eso que, además de la visualización, se pueden manejar diferentes técnicas estadísticas, como análisis de correspondencias [7], análisis de factores, entre otros.

El siguiente paso es realizar un proceso de **Modificación** de algunos datos mediante la creación, selección y transformación de variables que requieran un mejor enfoque para el modelado, tomando en cuenta los diagramas de Pareto [8] asociados a las mismas.

Concluida la fase anterior, se realiza el **Modelado** de la información mediante la hibridación de técnicas de minería de datos.

Fase 2. Segmentación. En ella se clasifican los clientes del producto bancario de acuerdo con el saldo de sus cuentas (Figura 10) y otras variables clave -estadísticamente-, para así poder identificar los subuniversos de cada segmento; mediante la técnica de segmentación, basada en clustering -una aplicación del vecino más cercano-, que se divide de forma proporcional al rango de saldos donde se concentra el mayor porcentaje de la población, a fin de no considerar saldos atípicos.

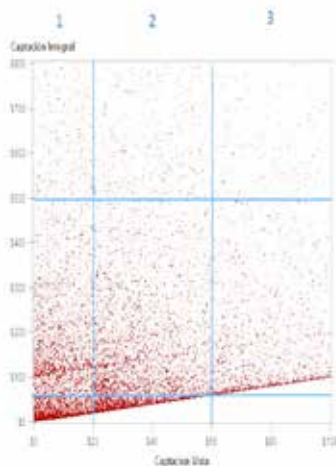


Figura 10

Segmentación de cuentas con base a sus saldos.

Fase 3. Análisis Multivariado. Mediante el diagrama de Pareto (Figura 11) se identifican las características demográficas, geográficas y psicográficas con mayor frecuencia en cada grupo, para conocer el perfil de los cuentahabientes por cada segmento.

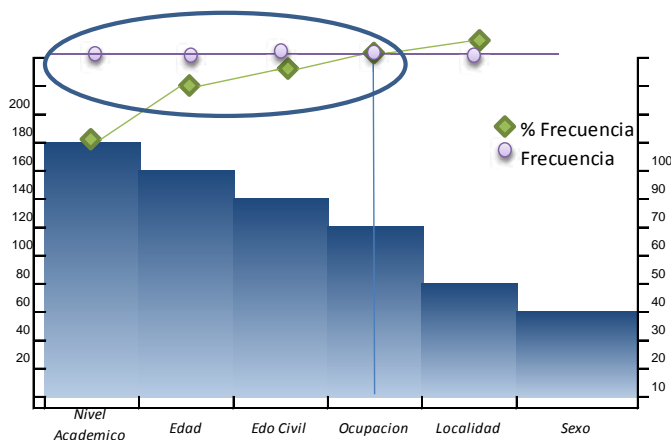


Figura 11.

Diagrama de Pareto con principales características de los cuentahabientes bancarios.

Fase 4. Regresión. Se analiza por segmento cada uno de los productos que ha adquirido el cuentahabiente a lo largo de su estancia y que han ayudado a ser rentables para la institución y así pronosticar el crecimiento que tendrá el segmento y el tiempo que tardará en alcanzar el siguiente nivel, estimando el crecimiento con las técnicas de venta que se tienen actualmente (Figura 12).

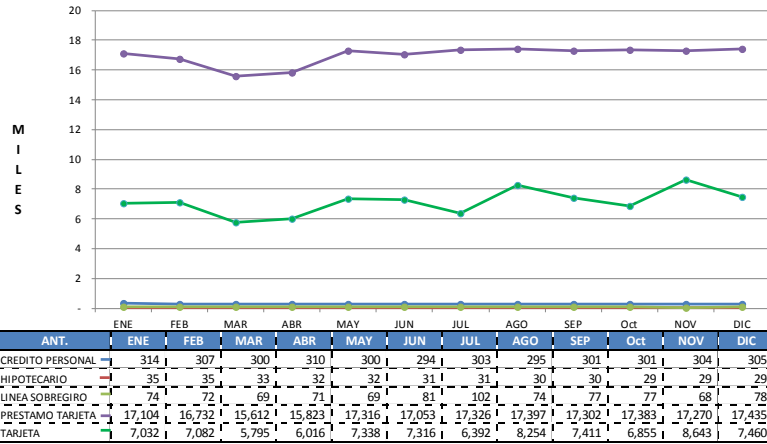


Figura 12

Productos que han adquirido los cuentahabientes a lo largo de su estancia en la institución.

Fase 5. Implementación. Consiste en colocar en ambiente de producción el modelo desarrollado, validando que el resultado arrojado sea el mismo al obtenido con el proceso anterior (Figura 13). El modelo permitirá conocer los segmentos en los que se divide el producto y el perfil de los cuentahabientes.

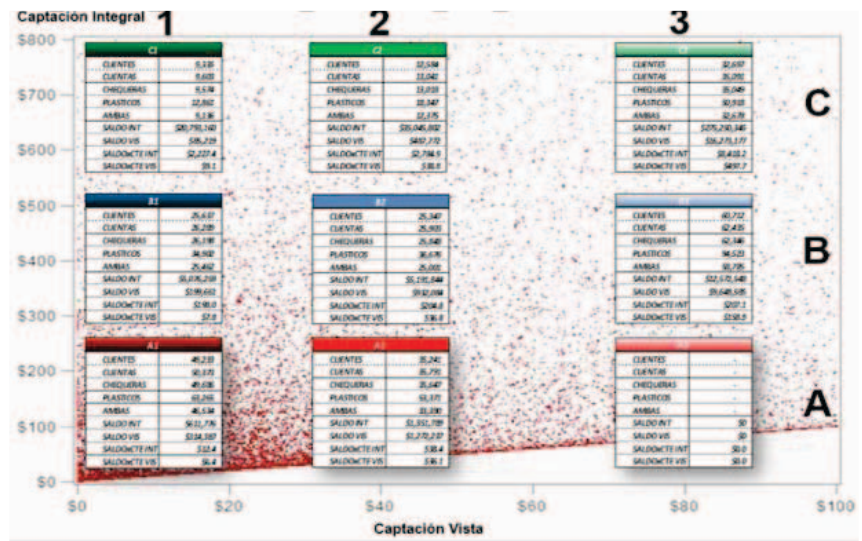


Figura 13

Interpretación del modelo analítico.

Fase 6. Resultado. Finalmente, como resultado de este modelo, se tiene la interpretación gráfica de las tendencias de rentabilidad, considerando para ello el nivel de transaccionalidad del cuentahabiente por cada segmento del mercado para cada producto (Figura 14), así como la disminución de tiempo de generación de métricas.

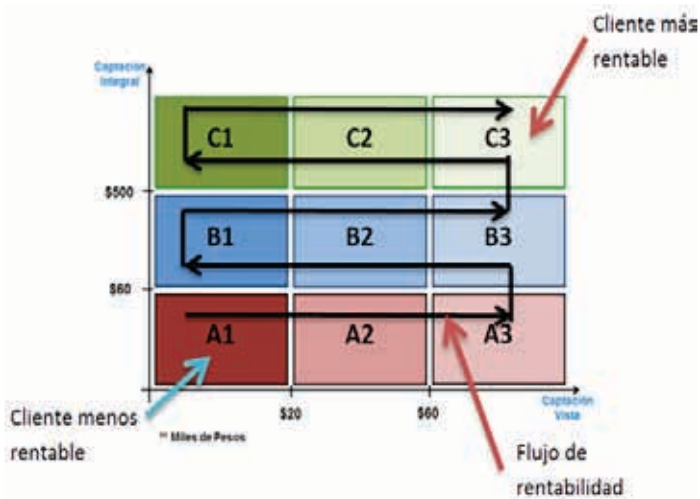


Figura 114

Clasificación de los clientes y flujo la rentabilidad.

Conclusiones

La implementación de un modelo que fusiona la técnica de minería de datos, permite identificar patrones ocultos en información histórica de los cuentahabientes bancarios, permitiendo conocer sus hábitos reales, a fin de mejorar la toma de decisiones para el mercadeo, la venta de productos y la mejora de servicios basados en la realidad de los mercados.

El modelo propuesto permite medir las tendencias en saldo integral (saldo de todas sus cuentas débito e inversiones) y saldo a la vista (saldo de la cuenta) dado cierto producto bancario y tomando como base los saldos de periodos anteriores mediante regresiones.

También permite determinar las variables de mayor impacto en los cuentahabientes, lo que facilita determinar qué tipos de productos son más atractivos para los mismos, así como estimar el tiempo en que un cuentahabiente podrá alcanzar el siguiente nivel de rentabilidad.

Se propone la implementación del modelo y la validación de los resultados arrojados con datos reales.

Asimismo, implementar el modelo utilizando la herramienta de SAS Enterprise Guide y Enterprise Miner, para tener mayor flexibilidad en las entregas y pruebas en cada proceso, reduciendo el riesgo de resultados incorrectos en el momento de la entrega final.

Se recomienda el uso modelos dinámicos adaptativos y agentes inteligentes que modifiquen el modelo a partir del perfil del cuentahabiente.

Referencias bibliográficas

- [1] Macías Rodríguez, M. (2016). *Técnicas de Minería de Datos para la Retención de Clientes en el Sector Asegurador*; http://www.cnsf.gob.mx/Eventos/Premios_2014/ANIVDELAREV.pdf (2008), Accedido el 2 de Febrero de 2016.
- [2] Díaz Diez, B., y Morillas Raya, A. (2016). *Minería de datos y lógica difusa. Una aplicación al estudio de la rentabilidad económica de las empresas agroalimentarias en Andalucía*; https://www.google.com.mx/?gfe_rd=cr&ei=Ev4vV_fLGLD18gfG7ZfwCQ&gws_rd=ssl#q=mineria+de+datos+aplicaciones&start=40
- [3] Benítez, I. J. (2005). *Técnicas de Agrupamiento para el Análisis de Datos Cuantitativos y Cualitativos*. Obtenido del trabajo de investigación publicado por la Universidad Politécnica de Valencia: Departamento de Ingeniería de Sistemas y Automática.
- [4] Pérez, C. (2009). *Técnicas Estadísticas Multivariantes con SPSS*. Madrid: Garceta grupo editorial.
- [5] Pereira, A. (2010). *Análisis Predictivo de Datos Mediante Técnicas de Regresión Estadística*. Universidad Complutense de Madrid.
- [6] Sobral, E. M. (2014). *Estudios elementales de legislación bancaria*, México. Obtenido de Información publicada en portal de Asociación de bancos de México: <https://www.abm.org.mx/la-banca-en-mexico/abm-historia.htm>
- [7] Turrent, E. (2015). *Historia sintética de la Banca en México*. Obtenido de Banco de México: file:///D:/1%20Maestria/2-Segundo%20Semestre/Seminario%20II/TESIS-ARTICULOS/Antecedentes/3_Banco%20de%20Mexico.pdf
- [8] Martínez, A. G. (2015). *La Banca en México*. Obtenido de Banco de México: <http://www.banxico.org.mx/>.
- [9] Vena, P. (2014). *Detección de datos atípicos para datos funcionales asimétricos*. Universidad de Buenos Aires, Facultad de Ciencias Naturales y Exactas.
- [10] De la Fuente, S. (2011). *Análisis Correspondencias Simples y Múltiples*. Universidad Autónoma de Madrid, Facultad de Ciencias Económicas y Administrativas.
- [11]. UNIT. (2009). *Herramientas para la mejora de la calidad*. Pearson, Instituto Uruguayo de Normas Técnicas.