

Información Importante

La Universidad de La Sabana informa que el(los) autor(es) ha(n) autorizado a usuarios internos y externos de la institución a consultar el contenido de este documento a través del Catálogo en línea de la Biblioteca y el Repositorio Institucional en la página Web de la Biblioteca, así como en las redes de información del país y del exterior con las cuales tenga convenio la Universidad de La Sabana.

Se permite la consulta a los usuarios interesados en el contenido de este documento para todos los usos que tengan finalidad académica, nunca para usos comerciales, siempre y cuando mediante la correspondiente cita bibliográfica se le de crédito al documento y a su autor.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, La Universidad de La Sabana informa que los derechos sobre los documentos son propiedad de los autores y tienen sobre su obra, entre otros, los derechos morales a que hacen referencia los mencionados artículos.

BIBLIOTECA OCTAVIO ARIZMENDI POSADA
UNIVERSIDAD DE LA SABANA
Chía - Cundinamarca

APLICACIÓN DE UN MODELO DE PREDICCIÓN DE FUGA DE CLIENTES EN LA MESA DE DINERO DE UNA ENTIDAD BANCARIA

Daniel Felipe Ruiz Londoño

Universidad de La Sabana

EICEA

2015

RESUMEN

En este trabajo se aplica un modelo de predicción de fuga de clientes con la información de los clientes de la mesa de dinero de una entidad bancaria en Colombia. Se busca analizar la información relevante que se debe tener en cuenta para este tipo de negocios no solo de manera estadística sino haciendo uso también de la intuición económica. Es por ello que en este trabajo se analizan variables económicas y financieras que son exógenas a la información de los clientes. El estudio se realiza con la aplicación de un modelo de regresión logística para obtener la probabilidad de fuga de los clientes en un periodo futuro. Se analizan algunos problemas con las variables para la aplicación de este modelo. Los resultados muestran un nivel de predicción aceptable para la consolidación de una estrategia comercial efectiva.

Tabla de contenido

I. INTRODUCCIÓN.....	3
II. REVISIÓN DE LITERATURA	4
III. DATOS	6
IV. MARCO TEÓRICO	10
V. METODOLOGÍA EMPÍRICA.....	12
VI. ESTIMACIONES Y RESULTADOS.....	16
VII. CONCLUSIONES.....	20
Agradecimientos	21
BIBLIOGRAFÍA.....	22

I. INTRODUCCIÓN

La gestión comercial de clientes de la mesa de dinero de un banco es de vital importancia para el desarrollo de las actividades del banco, así como para el desarrollo y eficiencia del mercado de valores en Colombia. Esta función es ejercida por la mesa de distribución, específicamente por los traders de distribución, quienes tienen a su cargo las relaciones comerciales con los clientes. La correcta ejecución de la estrategia comercial dependerá de la capacidad de análisis de la información que proveen los clientes, así como de las variables de los mercados financieros, además de la identificación de las necesidades y patrones de comportamiento de los clientes. Para ello es necesario desarrollar herramientas que permitan ofrecer un portafolio integral de productos bancarios que cubran las necesidades de ahorro y apalancamiento de las empresas, así como las de cobertura e inversión.

Una relación comercial exitosa debe permanecer durante un periodo prolongado en el tiempo, es por ello que se deben generar indicadores de gestión de la relación con los clientes, de tal manera que se puedan mitigar los riesgos de perder una relación comercial. Es por ello, que el objetivo de este trabajo es generar un modelo de probabilidad de fuga de clientes que funcione como indicador de alerta para los traders con el fin de fortalecer su relación comercial antes de que la fuga se haga efectiva.

Existen varios trabajos empíricos sobre este tipo de modelos en la industria de las telecomunicaciones y algunos en productos de crédito de banca privada. Sin embargo, no existen trabajos que estudien el caso de la mesa de dinero de una entidad bancaria, en donde además de las clásicas variables de negocio y de servicio al cliente, se deben tener en cuenta las variables de mercado que afectan la demanda de productos de tesorería. Ampliar la literatura sobre este tema es el principal aporte de este trabajo. Al final se logra predecir hasta

el 64% de las fugas para el periodo estudiado.

Este trabajo se organiza de la siguiente manera. En la primera parte se hace una revisión de literatura sobre el tema. En la siguiente sección se explican los datos a utilizar dentro del modelo propuesto, luego se detalla el modelo a utilizar y la metodología empírica. Para finalizar se exponen los resultados obtenidos y las conclusiones finales del estudio.

II. REVISIÓN DE LITERATURA

Hoy día, la minería de datos está cobrando relevancia creciente en empresas y organizaciones para resolver problemas complejos de negocio, basado en el procesamiento de volúmenes masivos de datos donde se esconde información valiosa respecto del comportamiento de compra de productos o servicios (Barrientos 2013). El estudio de esta información puede llegar a ser bastante complejo dada la cantidad de información, lo que lo convierte en un trabajo interdisciplinar que recoge herramientas de marketing, ingeniería y economía con el fin de poder automatizar los procesos de análisis de comportamiento de clientes. Algunas de las aplicaciones de la minería de datos que hacen uso de las áreas del conocimiento mencionadas anteriormente son los modelos de predicción de fuga de clientes en una empresa o “churn” como se encuentra en la literatura científica.

Según (Gür Ali and Aritürk 2014), existe bastante literatura en cuanto a modelos y metodologías de predicción de fuga de clientes. En un inicio la literatura se enfocó en la clasificación de metodologías en términos de la precisión en su predicción, tales como regresiones logísticas, árboles de decisión, redes neuronales, análisis de supervivencia, etc. Sin embargo, ahora los investigadores estudian el impacto de los datos, su procesamiento, el

muestreo y la representación de los datos de los clientes en la precisión de la predicción, independientemente de la metodología o algoritmo de clasificación que se use. Además, autores como (Ahn, Han, and Lee 2006; Echeverría and Retamal 2011; Gür Ali and Aritürk 2014; López 2013; Nie et al. 2011; Segovia, Aburto, and Goic 2000) han realizado trabajos empíricos con las metodologías revisadas en la literatura científica, aplicándolas en empresas reales, especialmente de servicios tecnológicos como empresas de telecomunicaciones, telefonía celular, productos de consumo o productos financieros básicos.

Si bien existen varios trabajos teóricos y empíricos sobre el tema de esta investigación, ninguno de los autores, a pesar de aplicar las metodologías en instituciones financieras y productos financieros, llegan a tener en cuenta dentro de sus variables explicativas o determinantes, algún tipo de variable económica o financiera exógena que sea ajena a la información obtenida por la operación de los clientes con las empresas, tales como la volatilidad de la tasa de cambio, indicadores financieros de la empresa, inflación, tasa de política monetaria, etc. Estas variables son determinantes clave en la demanda de productos financieros de tesorería por parte de las empresas del sector real de la economía tales como forwards, swaps, opciones, spot, etc., como podemos encontrar en autores como (Bodnar, Marston, and Hayt 1998; Martin et al. 2009; Pozo-Olano 1974).

El presente estudio busca por tanto, incluir en los modelos actualmente utilizados en la predicción de fuga de clientes, variables de la economía que afectan el comportamiento de los clientes de este tipo de productos financieros, y al mismo tiempo proponer una herramienta que por tratarse de un área muy técnica y financiera, deje de lado temas de marketing que bajo metodologías de análisis estadístico y económico del negocio, puede proveer información importante y consistente para la toma de decisiones de negocio que se

pueden transformar en mayores utilidades en el mediano plazo, utilizando las estrategias adecuadas. Esto haciendo uso de la información que provee el negocio propio y el contexto económico actual, lo cual la convierte en una herramienta de bajo de costo con capacidad de generar valor para una entidad bancaria.

III. DATOS

Para la generación del modelo de predicción de fuga de clientes se obtienen los datos de clientes y operaciones de la tesorería de la entidad bancaria de los sistemas de registro de operaciones y gestión de portafolios internos. Con el fin de extraer la información necesaria para la creación de la base de datos de este estudio fue necesaria la extracción de la información de tres sistemas de registro diferentes que varían según el tipo de producto de la tesorería, y crear un formato estándar para su uso integrado. Los datos aquí obtenidos se encuentran en un rango de fechas entre Enero de 2011 hasta Diciembre de 2014. El objetivo es utilizar la información histórica de operaciones por cliente y extraer patrones de comportamiento que permitan calibrar el modelo de regresión logística que se propone en esta investigación, para así luego probar su capacidad predictiva.

De esta manera entonces se obtiene una base de datos de los clientes de la tesorería con las variables relevantes para este estudio. Si bien existen trabajos relacionados con este tema, la selección de variables y construcción de los datos varía dependiendo del tipo de negocio de las empresas en las que se han aplicado este tipo de modelos. De acuerdo a la revisión de la literatura, no se encuentran referencias de estudios relacionados con productos de tesorería en una entidad financiera.

Para propósitos del análisis, un cliente fugado se definió como aquel que habiendo operado en uno o más periodos anteriores (año), no realizó operaciones en el periodo actual. Basado en esta definición, también se definieron otras variables categóricas que representan el comportamiento del cliente en cuanto a consumo de productos de tesorería, como clientes fieles, los cuales se definieron como clientes que operaron durante todos los periodos analizados. Así en total se obtuvo una base de datos con información completa de 70000 clientes, que incluye las siguientes variables: montos operados en dólares, número de operaciones, productos operados, otros productos que se tienen con el banco, utilidad generada para la tesorería, y número de traders diferentes que han atendido. Con el fin de calibrar el modelo con la información de clientes que realmente son importantes para el negocio, se filtró la base con solo los clientes carterizados, es decir, clientes a los que el banco les ha asignado un ejecutivo de cuenta para mantener una mejor relación comercial, quedando así un total de 3097 clientes para analizar. Cabe resaltar que esta información fue comparada con los informes de gestión presentados en cada periodo analizado. Sumado a esta información, y para fines de un correcto análisis dadas las características del negocio, se añadieron variables económicas y financieras que son determinantes en la demanda de productos de tesorería y que no han sido tenidas en cuenta en otros estudios relacionados tales como inflación, volatilidad de tasa de cambio, índice de volatilidad de mercado, tasa repo y tasas de captación de entidades financieras.

Por último, es necesario dar a conocer una posible limitación de los datos que podría afectar los resultados del modelo. Para ello, primero se debe aclarar que un supuesto de este estudio es que los clientes que adquieren productos de tesorería con el banco, tienen algún tipo de actividad económica en moneda extranjera, ya sea por actividades de importación o

exportación o por la realización de giros en el exterior. Para identificar esta información, se obtuvieron datos de las declaraciones de importaciones y exportaciones en Colombia durante los periodos de análisis de este estudio directamente de la DIAN (es información pública) y se consolidó una base de datos de alrededor de 80000 empresas en donde se identifica si la empresa es exportador neto (sólo realiza actividades de exportación), importador neto (sólo realiza actividades de importación), mixta (realiza actividades tanto de importación como de exportación) u ocasional (han realizado algún actividad de importación y exportación ocasionalmente). Además, también se identifica si para cada categoría mencionada, el cliente tiene interrupciones (no aparecen datos en algún periodo analizado) o constante (tiene datos en todos los periodos), en donde se encontraron alrededor de 54000 y 16000 observaciones respectivamente. Al cruzar esta base de datos de la DIAN con la base de clientes de la tesorería, el número de observaciones con información completa se reduce drásticamente hasta alrededor de 5000. Esto puede llevar a una lectura errónea de los datos, pues para una empresa sin información en la DIAN que puede ser identificada como fugada pueden omitirse otros factores que generen “fuga” como la quiebra. Sin embargo, se generó una dummy que toma el valor de 1 si existe información del cliente en las bases de DIAN para un año determinado o 0 si no para motivos de análisis del modelo.

- Descripción de Variables

Tabla 2 - Total variables según tipo	
<i>Total Variables dummy</i>	9
<i>Total Variables Categóricas</i>	1
<i>Total Variables Continuas</i>	13
<i>Total Variables</i>	23

Fuente: Elaboración Propia

Nombre de la Variable	Descripción	Tipo de Variable
<i>Antigüedad del Cliente</i>	Es una variable proxy del valor de la antigüedad del cliente.	Continúa
<i>Combos</i>	Variable dummy que determina si el cliente tiene combos de portales con el banco	Dummy
<i>DIAN</i>	Variable dummy que determina si el cliente reporta operaciones de exportaciones o importaciones ante DIAN	Dummy
<i>DTF</i>	Es la tasa promedio de los CDT's de los últimos 30 días	Continúa
<i>Forex</i>	Variable dummy que toma valor 1 si el cliente operó el producto Forex o 0 si no	Dummy
<i>Forward</i>	Variable dummy que toma valor 1 si el cliente operó el producto Forward o 0 si no	Dummy
<i>Forward-Forex</i>	Variable dummy que toma valor 1 si el cliente operó el producto Forward-Forex o 0 si no	Dummy
<i>IPC</i>	Es el índice de precios al consumidor de la economía en un periodo determinado	Continúa
<i>Monto Operado</i>	Es el monto operado por el cliente en un periodo determinado	Continúa
<i>Next Day</i>	Variable dummy que toma valor 1 si el cliente operó el producto Next Day o 0 si no	Dummy
<i>Número de operaciones</i>	Es el número de operaciones realizadas por el cliente en un periodo determinado	Continúa
<i>Opciones</i>	Variable dummy que toma valor 1 si el cliente operó el producto Opciones o 0 si no	Dummy
<i>Portales</i>	Variable dummy que determina si el cliente tiene el servicio de portales con el banco	Dummy
<i>Spot</i>	Variable dummy que toma valor 1 si el cliente operó el producto Spot o 0 si no	Dummy
<i>Segmento</i>	Variable categórica que indica el segmento al que pertenece el cliente	Categórica
<i>Tasa Repo</i>	Muestra la tasa de intervención del Banco de la República, como proxy de los costos de fondeo.	Continúa
<i>Traders</i>	Es el número de traders diferentes con los que un cliente operó en un periodo de tiempo determinado	Continúa
<i>Utilidad</i>	Es la utilidad que obtuvo la tesorería sobre las operaciones de un cliente determinado en un periodo determinado	Continúa
<i>VIX</i>	Índice de Volatilidad del CBOE	Continúa
<i>Volatilidad USDCOP</i>	Mide la volatilidad promedio del par USDCOP en un periodo determinado	Continúa
<i>Volatilidad EURUSD</i>	Mide la volatilidad promedio del par EURUSD en un periodo determinado	Continúa
<i>Volatilidad GBPUSD</i>	Mide la volatilidad promedio del par GBPUSD en un periodo determinado	Continúa
<i>Volatilidad USDJPY</i>	Mide la volatilidad promedio del par USDJPY en un periodo determinado	Continúa

Fuente: Elaboración Propia

Tabla 3 - Variables en el Modelo	
Nombre	Variable
<i>Antigüedad del Cliente</i>	antig
<i>Combos</i>	combos
<i>DIAN</i>	dian
<i>DTF</i>	dtf
<i>Forex</i>	forex
<i>Forward</i>	forward
<i>Forward-Forex</i>	forward_fx
<i>IPC</i>	ipc
<i>Monto Operado</i>	monto
<i>Next Day</i>	next_day
<i>Número de operaciones</i>	num_op
<i>Opciones</i>	opciones
<i>Portales</i>	portales
<i>Spot</i>	spot
<i>Tasa Repo</i>	repo
<i>Trader</i>	traders
<i>Utilidad</i>	util
<i>VIX</i>	vix
<i>Volatilidad EURUSD</i>	eurusd
<i>Volatilidad GBPUSD</i>	gbpusd
<i>Volatilidad USDCOP</i>	usdcop
<i>Volatilidad USDJPY</i>	usdjpy

Fuente: Elaboración Propia

Para construir la variable proxy del valor de la antigüedad del cliente, se toma como base el número de años que el cliente operó en el periodo analizado y este valor corresponderá al primer año. Con el fin de dar o quitar valor a la antigüedad, se comienza a sumar o restar una constante igual a $\frac{1}{c}$ donde c corresponde al número de años que se van a analizar. Así por ejemplo si para un periodo de 4 años a analizar un cliente operó en solo un año, este será el valor que tomara en el primer año de análisis; para el siguiente año, se revisa si el cliente operó en ese periodo, si es así, entonces al valor del año anterior se le

sumaría 0.25, así esta variable toma un valor de 1.25 para el segundo año analizado, en caso

de que no haya operado entonces la variable toma un valor de 0.75.

Las demás variables se obtienen de las bases de datos de clientes y de sistemas de información financiera.

IV. MARCO TEÓRICO

La regresión logística, la cual es muy utilizada como técnica de modelamiento estadístico, puede ser utilizada para construir un modelo de respuesta binaria o dicótoma que surge de los modelos de probabilidad lineal, en donde a la variable dependiente de respuesta bivariada también se le conoce como la probabilidad de respuesta (Nie et al. 2011). El modelo de regresión logística es un procedimiento por medio del cual se analizan las relaciones de asociación entre una variable dependiente binaria y una o varias variables independientes cuantitativas o categóricas, con el fin de lograr determinar la existencia o ausencia de relación entre una o más variables independientes y la variable dependiente, medir la magnitud de dicha relación y estimar o predecir la probabilidad de que se produzca el suceso definido por la variable dependiente en función de los valores que adoptan las variables independientes (Echeverría and Retamal 2011). La regresión logística ha sido muy estudiada y usada en diversas aplicaciones como la agricultura, el sobrepeso y la obesidad, credit scoring, análisis de accidentes y prevención (Nie et al. 2011).

La forma específica del modelo de regresión logística es:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1)$$

La transformación de la función logística $\pi(x)$ es conocida como la transformación logit:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x \quad (2)$$

El método tradicional de estimación que lleva a la función de mínimos cuadrados bajo el modelo de regresión lineal es el método de máxima verosimilitud el cual provee las bases para estimar los parámetros del modelo de regresión logística. Este trabajo se basa en este modelo conocido y expuesto en otros trabajos relacionados, en vez de utilizar uno alternativo (Nie et al. 2011).

Uno de los métodos para probar la significancia conjunta de las variables incluidas en un modelo de regresión logística, es a través de la prueba de Wald, que requiere la estimación solo del modelo no restringido. En el caso del modelo lineal, el estadístico de Wald después de una simple transformación, es esencialmente el estadístico F. Éste tiene una distribución ji-cuadrada asintótica, con grados de libertad igual al número de restricciones que se están probando (Wooldridge 2009). Si el valor experimental de Wald, bajo los supuestos de la hipótesis nula ($H_0: \beta_k = 0, H_1: \beta_k \neq 0$), es menor que el nivel de significancia fijado en el contraste, se rechaza la hipótesis nula y se concluye que las variables independientes influyen en la probabilidad de las características de la variable dependiente (Echeverría and Retamal 2011).

Una de las limitaciones de este modelo es que una potencial falta de independencia introducida por observaciones múltiples de un mismo cliente pueda sesgar las estimaciones de los parámetros y/o incrementar artificialmente la significancia de los parámetros con métodos que asumen errores independientes e idénticamente distribuidos, como la regresión logística (Gür Ali and Aritürk 2014).

V. METODOLOGÍA EMPÍRICA

Dado que el objetivo de este trabajo es obtener un modelo que permita obtener la probabilidad de fuga de un cliente determinado en un periodo posterior, necesitamos un modelo cuya variable dependiente sea una variable de probabilidad de respuesta tal y como se mencionó en el marco teórico. Es por ello que para el desarrollo de este trabajo se estimarán los parámetros del modelo a través del método de regresión logística.

Muchos economistas prefieren el modelo probit sobre el modelo logit para estimar modelos de probabilidad de respuesta. Esta preferencia se debe a que el modelo probit asume una distribución normal de los datos para generar un error que se distribuye normalmente con lo que es más fácil de trabajar. Dada la diversidad de clientes que se encuentran en la base de datos construida, y que su comportamiento difiere dependiendo del sector y segmento en el que se clasifica, es difícil cumplir con la normalidad de los datos, por lo cual para efectos de simplicidad y evitar la normalización de los datos, se decide emplear el modelo logit o de regresión logística, propuesto en otros trabajos relacionados como (Barrientos 2013; Echeverría and Retamal 2011; Grado n.d.; Gür Ali and Aritürk 2014; Nie et al. 2011).

Para entender la intuición que se trata de modelar en este trabajo, es necesario entender el tipo de clientes que se tienen en una mesa de dinero y la función de esta área.

La mesa de dinero de una tesorería de un banco, entidad financiera o empresa del sector real, es el área encargada de gestionar los defectos del activo y el pasivo, e invertir o fondear los excesos o faltantes de liquidez para el correcto funcionamiento de la organización a través de operaciones en el mercado de valores, según la coyuntura. Las mesas de dinero tienen diferentes divisiones según el mercado al que afecten. Normalmente se encuentran divididas

en varias mesas como mesa de divisas, mesa de moneda legal, mesa de deuda pública y mesa corporativa. La mesa corporativa es también conocida como mesa de distribución, porque es allí donde se atiende a los clientes y se distribuyen los productos del resto de las mesas. Es por ello que este trabajo se basa en la operación de la mesa de distribución.

Los clientes que adquieren los productos de la mesa de dinero en un banco, pueden ser personas naturales (con montos mínimos de operación) y empresas del sector real que son segmentadas por el banco según un criterio específico. Por ejemplo, se organizan por segmentos pyme o corporativo. Para efectos de este trabajo se dejan por fuera los clientes institucionales que operan con la mesa de dinero del banco por tener un comportamiento diferente. La motivación principal para que los clientes estudiados adquieran productos en la mesa de dinero es el hecho de tener algún tipo de operación en moneda extranjera, ya sea un giro al exterior, una monetización, o una cobertura de un flujo en moneda extranjera en una fecha futura. Para este efecto, adquieren productos de la mesa de divisas como spot (compra/venta de moneda extranjera a la tasa spot), forwards y futuros (compra/venta de moneda extranjera en una fecha futura) u opciones con fines de cobertura ante situaciones económicas de alta volatilidad en la tasa de cambio. Normalmente las personas jurídicas que operan con la mesa de dinero tienen actividades de exportación o importación de productos.

Los traders de la mesa de distribución son los encargados de tener el contacto directo con los clientes y asesorarlos en los productos que deben adquirir según la naturaleza de su negocio y la coyuntura económica del momento. Para ello se establece una estrategia comercial que consiste en continuas visitas presenciales o llamadas telefónicas con el fin de mantener excelentes relaciones comerciales con los clientes. Sin embargo, la cantidad de clientes encontrados en la base de datos en proporción con la cantidad de traders supone un gran

problema para una correcta gestión de clientes y de esta manera garantizar una excelente experiencia de servicio. Por este motivo es difícil establecer un criterio suficiente de gestión para evitar la alta rotación o fuga de clientes de un periodo a otro teniendo en cuenta el costo/beneficio de dicha estrategia. Es por ello que el objetivo de este trabajo es generar un criterio de gestión de clientes basado en la probabilidad de fuga de un cliente en determinado periodo, basado en su comportamiento histórico y actual.

Dadas las diversas variables exógenas que afectan la decisión de un cliente para adquirir un producto de tesorería que incluyen variables económicas entre otras, que es el aporte de este trabajo a la literatura, entonces se incluyen las siguientes variables en el modelo:

- *portales* y *combos* son variables dummy que identifican clientes que tienen estos servicios con el banco, y que por ende es más factible que también adquieran productos de tesorería con el mismo banco dado la dificultad y tramitología que fue necesaria para adquirir estos servicios y los beneficios que suponen. Se refieren a otros productos del banco que no se gestionan desde la mesa de dinero sino a través de los gerentes de cuenta.
- Las demás variables incluidas en el modelo se detallaron en la **Tabla 1** de la sección de datos. Sin embargo, vale la pena explicar la intuición detrás de la inclusión de variables económicas dentro del modelo tales como *dtf*, *ipc*, *repo*, *vix*, *eurusd*, *gbpusd*, *usdcop*, *usdjpy*. Como se mencionaba anteriormente, la motivación de los clientes para adquirir productos de tesorería generalmente es la necesidad de realizar operaciones en moneda extranjera. Como sabemos de la teoría económica, la tasa de cambio se mueve ante diversos choques en variables macroeconómicas como las

tasas de interés de la economía (tasa repo), la cual a su vez puede depender por lo menos en el caso colombiano de la tasa de inflación de la economía (ipc). Además, la tasa de cambio también depende del flujo de capitales extranjeros en la economía, los cuales en países como Colombia se dan en situaciones de apetito por el riesgo “search for yield” que provocan revaluaciones en la moneda local o por el contrario la salida de capitales por falta de apetito por riesgo de mercados emergentes “fly to quality” que provoca devaluaciones en la moneda local. Una variable que mide el pánico en los mercados y por tanto es una proxy de este tipo de comportamientos es el índice de volatilidad VIX. Esta variable además permite incluir las volatilidades de las tasas de cambio de las monedas más operadas en la mesa de dinero. Cabe anotar que dichas variables se toman como promedios mensuales de cada periodo i -ésimo medido en años.

- Se incluye una variable proxy de percepción del servicio proporcionado por el trader. La variable *traders* indica la cantidad de traders diferentes con los que operó un cliente determinado en el periodo i -ésimo y estimar de esta manera la importancia que le da el cliente al vínculo que se genere con un trader. Actualmente es normal que los clientes sean atendidos por varios traders debido a la operativa de la mesa de dinero de la entidad en estudio, sin embargo, la significancia económica de esta variable sería un motivo suficiente para establecer nuevas políticas de gestión de clientes.
- La variable *antig* es una variable proxy de la antigüedad del cliente a través del periodo tenido en cuenta para entrenar el modelo al tener en cuenta el número de años

que operó y un puntaje por la constancia de la operación con el banco año a año.

- Por último, dado que los clientes se encuentran clasificados en diferentes segmentos, se incluye una variable categórica para diferenciar entre los segmentos de cada cliente descritos en la **Tabla 6**.
- La base de datos conformada por las variables explicadas es una base de datos construida con la información obtenida de los sistemas de gestión de portafolios de la tesorería. Las variables categóricas se construyeron con fórmulas condicionales según los criterios establecidos.

VI. ESTIMACIONES Y RESULTADOS

Una vez construida la base de datos en su totalidad, se procedió a realizar el análisis de los datos y las estimaciones del modelo Logit propuesto para la predicción de fuga de clientes.

Los resultados obtenidos se presentan a continuación.

Con el fin de conocer el comportamiento de los clientes de manera general y teniendo en cuenta las salidas de la variable dependiente *fugados* cuyos

Tabla 4 - Fugados	
Estado	Indicador
<i>Fugado</i>	1
<i>Activo</i>	0

Fuente: Elaboración Propia

valores se especifican en la tabla 4, se procede a construir las matrices de probabilidad de transición en base a los datos históricos obteniendo así los resultados reportados en la tabla

5.

Tabla 5 - Matriz de Transición General			
Fugados	0	1	Total
0	76.46%	23.54%	100%
1	43.42%	56.58%	100%
Total	63.10%	36.90%	100%

Fuente: Elaboración Propia, cálculos en Stata

De esta manera, podemos observar que la probabilidad de que un cliente activo en el periodo Y_t permanezca activo en el periodo Y_{t+1} es del 76.46%, pero la probabilidad de que se convierta en fugado es del 23.54%. El mismo análisis se puede realizar con los clientes fugados, así un cliente fugado en el periodo Y_t tiene una probabilidad de convertirse de nuevo en cliente activo en el periodo Y_{t+1} del 43.42%, y una probabilidad del 56.58% de permanecer como fugado. Por último, se puede decir que para los periodos analizados, la tasa de clientes activos es del 63.10% mientras que la tasa de fuga es del 36.90%.

Tabla 6 - Segmentos	
Segmento	Indicador
<i>Institucional</i>	1
<i>Corporativo</i>	2
<i>Pyme</i>	3
<i>Barranquilla</i>	4
<i>Cali</i>	5
<i>Medellin</i>	6

Fuente: Elaboración Propia

Estas matrices también se construyen para cada segmento de clientes que se especifica en la tabla 6, y se interpretan de la misma manera. A continuación se presentan las matrices:

Tabla 7 - Matriz de Transición - 1			
Fugados	0	1	Total
0	82.57%	17.43%	100%
1	42.06%	56.58%	100%
Total	72.02%	27.98%	100%

Fuente: Elaboración Propia, cálculos en Stata

Tabla 9 - Matriz de Transición - 3			
Fugados	0	1	Total
0	73.47%	26.53%	100%
1	43.19%	56.81%	100%
Total	60.32%	39.68%	100%

Fuente: Elaboración Propia, cálculos en Stata

Tabla 8 - Matriz de Transición - 2			
Fugados	0	1	Total
0	82.11%	17.89%	100%
1	43.45%	56.55%	100%
Total	67.35%	32.65%	100%

Fuente: Elaboración Propia, cálculos en Stata

Tabla 10 - Matriz de Transición - 4			
Fugados	0	1	Total
0	76.81%	23.19%	100%
1	42.12%	57.88%	100%
Total	62.10%	37.90%	100%

Fuente: Elaboración Propia, cálculos en Stata

Fugados	0	1	Total
0	75.65%	24.35%	100%
1	44.29%	55.71%	100%
Total	62.89%	37.11%	100%

Fuente: Elaboración Propia, cálculos en Stata

Fugados	0	1	Total
0	75.37%	24.63%	100%
1	44.52%	55.48%	100%
Total	62.51%	37.49%	100%

Fuente: Elaboración Propia, cálculos en Stata

Una vez obtenidos estos resultados, se procedió a estimar el modelo Logit obteniendo los resultados presentados en la tabla 13 que se presenta a continuación:

Logistic regression		Number of obs	=	9289	
		LR chi2(7)	=	5798.80	
		Prob > chi2	=	0.0000	
Log likelihood = -3368.2182		Pseudo R2	=	0.4626	
fugados	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
antig	-1.700088	.0333108	-51.04	0.000	-1.765376 -1.6348
segmento	.0085868	.021707	0.40	0.692	-.0339581 .0511316
ipc	42.3049	4.187919	10.10	0.000	34.09673 50.51307
repo	10.84487	5.006311	2.17	0.030	1.032682 20.65706
vix	(omitted)				
portales	.3287218	.1008732	3.26	0.001	.1310139 .5264298
combos	.1038058	.1863656	0.56	0.578	-.2614641 .4690757
dian	-.0427934	.0627178	-0.68	0.495	-.1657181 .0801312
_cons	1.385117	.2521772	5.49	0.000	.8908587 1.879375

Fuente: Elaboración Propia, cálculos en Stata

Como podemos ver, no fue posible incluir todas las variables propuestas inicialmente en el modelo, debido a que variables como *monto*, *numop* o *útil*, explican perfectamente el modelo al tomar el mismo valor de 0 siempre que el cliente es fugado, y otras variables como el *vix* presentan problemas de colinearidad. Sin embargo, el modelo se corrió con las demás variables tenidas en cuenta tanto de mercado como de negocio, y aunque variables como *dian* no presentan significancia estadística per se, la prueba de verosimilitud a través del estadístico chi cuadrado, nos muestra que las variables presentan significancia conjunta, lo que justifica la inclusión de estas en el modelo.

Tabla 14 - Capacidad Predictiva del Modelo							
Probabilidad de Fuga	Predicciones de Fuga	Fugas Reales	% de predicción	Predicciones Verdaderas	Falsos Positivos	% de Predicción Real	% de Predicciones Falsas
>90%	159	482	32.99%	159	0	32.99%	0.00%
>80%	182	482	37.76%	182	0	37.76%	0.00%
>70%	182	482	37.76%	182	0	37.76%	0.00%
>60%	213	482	44.19%	213	0	44.19%	0.00%
>50%	217	482	45.02%	215	2	44.61%	0.41%
>40%	589	482	122.20%	312	277	64.73%	57.47%

Fuente: Elaboración Propia

Este modelo es el resultado de la calibración con los datos desde 2011 a 2013, con el fin de testear la capacidad predictiva del modelo con la información de 2014. Por ello se procedió a estimar las predicciones de fuga para el año 2014 con los resultados obtenidos de la estimación obteniendo los resultados presentados en la tabla 14 y que podemos interpretar de la siguiente manera:

Para clientes con estimaciones de probabilidad de fuga mayores al 90% en el 2014, se obtuvieron un total de 159 predicciones para ese periodo, en el cual se dieron realmente 482 casos de clientes fugados, teniendo así un poder de predicción del 32.99%. Del total de estas predicciones, las 159 correspondieron a clientes que realmente se fugaron en ese periodo por lo cual el modelo no arrojó falsos positivos (clientes con probabilidad de fuga alta que realmente no se fugaron en el periodo en cuestión), por lo cual el poder de predicción real del modelo es del 32.99%.

Para los demás rangos de probabilidad de fuga, la interpretación se realiza de la misma manera, sin embargo, es interesante recalcar que al analizar clientes con probabilidad de fuga mayores al 40%, si bien obtenemos un 57.47% de predicciones falsas, el modelo al mismo tiempo tiene un porcentaje de predicción real del 64.73%. Este resultado es bastante satisfactorio, pues es una oportunidad para reducir la tasa de fuga de un periodo a otro en más de un 50% lo cual tendría un impacto bastante positivo en los ingresos de la entidad.

En conclusión, el modelo arrojó resultados de impacto para la organización, logrando el objetivo planteado inicialmente de construir una herramienta que permita tener indicadores de referencia para plantear mejores estrategias de gestión comercial.

Con respecto a la interpretación de los coeficientes del modelo arrojados para cada variable, al ser un modelo logit, su interpretación no se puede realizar de manera directa. Para ello debemos obtener el odds ratio de cada variable dentro del modelo, el cual nos permitirá obtener el factor de multiplicación de la probabilidad de éxito sobre la probabilidad de falla de la variable dependiente. Así por ejemplo, el odds ratio de la variable antig para este modelo es equivalente a 0.1826675, lo que significa que la probabilidad de fuga sobre la probabilidad de ser activo disminuye en este factor cuando la variable antig aumenta en una unidad. Sin embargo, el objetivo de este trabajo no es el de analizar los efectos marginales de las variables, por lo que esta explicación es solo ilustrativa.

VII. CONCLUSIONES

En esta investigación se ha propuesto la aplicación de un modelo de predicción de fuga de clientes para la mesa de dinero de una entidad bancaria en Colombia. El propósito del trabajo no es el de generar un nuevo algoritmo o modelo, sino que se enfoca en la ejecución y entendimiento de un modelo propuesto en la literatura. Las variables y la extracción de la información dependerán del tipo de negocio que se quiera analizar.

La capacidad predictiva del modelo es aceptable bajo las variables que se pudieron incluir para su calibración a pesar de que la mayoría de estas no pudieron ser incluidas debido a problemas de convergencia o colinearidad dentro del modelo. Sin embargo, las variables que

se tomaron en cuenta demuestran que para este tipo de negocio es importante considerar variables del mercado, variables de los clientes y variables de servicio.

Para trabajos futuros, sería interesante trabajar las variables de los clientes como montos y números de operaciones que no pudieron ser incluidas en el modelo de tal forma que estas puedan incluirse y así mejorar el performance del modelo al incorporar el comportamiento de operación de los clientes.

Agradecimientos

Se presentan agradecimientos al banco anónimo del cual se obtuvieron los datos necesarios para realizar el análisis, así como el seguimiento sobre las variables importantes del negocio. También a la profesora Silvia Gómez Soler por su revisión y acompañamiento en la realización del trabajo.

BIBLIOGRAFÍA

- Ahn, Jae Hyeon, Sang Pil Han, and Yung Seop Lee. 2006. "Customer Churn Analysis: Churn Determinants and Mediation Effects of Partial Defection in the Korean Mobile Telecommunications Service Industry." *Telecommunications Policy* 30(10-11):552–68.
- Barrientos, Francisco. 2013. "Aplicación de Minería de Datos Para Predecir Fuga de Clientes En La Industria de Las Telecomunicaciones." 73–108.
- Bodnar, Gordon M., Richard C. Marston, and Greg Hayt. 1998. "Survey of Financial Risk Management by U . S . Non-Financial Firms Written." *Risk Management* 31.
- Echeverría, Mauricio E. and Víctor F. J. Ovalle Retamal. 2011. "APLICACIÓN DE UN MODELO PREDICTIVO DE FUGA DE CLIENTES UTILIZANDO DATA MINING EN VTR." *Tesis de Grado - Ingeniería Civil Industrial*.
- Grado, Trabajo Fin De. n.d. "Facultad de Ciencias."
- Gür Ali, Özden and Umut Aritürk. 2014. "Dynamic Churn Prediction Framework with More Effective Use of Rare Event Data: The Case of Private Banking." *Expert Systems with Applications* 41(17):7889–7903.
- López, Leonardo. 2013. "PREDICCIÓN DE FUGA DE CLIENTES DESDE UN ENFOQUE DE COMPETENCIA." *Tesis de Grado - Maestría en Gestión de Operaciones*.
- Martin, Miguel Ángel, Wolfgang Rojas, José Luis Eráusquin, Dayana Yupanqui, and Édgar Vera. 2009. "Derivatives Usage By Non-Financial Firms in Emerging Markets : The Peruvian Case." *Journal of Economics, Finance and Administrative Science* 14(27):73–86. Retrieved (http://www.esan.edu.pe/publicaciones/2009/12/07/journal_of_economics_finance_and_administrative_science/5-martin.pdf).
- Nie, Guangli, Wei Rowe, Lingling Zhang, Yingjie Tian, and Yong Shi. 2011. "Credit Card Churn Forecasting by Logistic Regression and Decision Tree." *Expert Systems with Applications* 38(12):15273–85. Retrieved (<http://dx.doi.org/10.1016/j.eswa.2011.06.028>).
- Pozo-Olano, J. D. D. 1974. "International Evidence on the Determinants of Foreign Exchange Rate Exposure of Multinational Corporations." *Science* 183(4124):469–70.
- Segovia, Carolina, Luis Aburto, and Marcel Goic. 2000. "Caracterización Del Procesos de Fuga de Clientes Utilizando Información Transaccional." (1998).
- Wooldridge, Jeffrey M. 2009. *Introducción a La Econometría: Un Enfoque Moderno*.