

# UCUENCA

Facultad de Ciencias Químicas

Carrera de Ingeniería Industrial

Modelo de segmentación de cartera basado en el comportamiento de clientes para la gestión de cobranza

Trabajo de titulación previo a la obtención del título de Ingeniera Industrial

Autora:

Michelle Estefanía Cabrera Calderón

CI: 0104638416

Correo electrónico: michellecab.cm@gmail.com

Tutor:

James Marlon Arias Cisneros

CI: 0102407145

**Cuenca, Ecuador**

21-julio-2022

## “Modelo de segmentación de cartera basado en el comportamiento de clientes para la gestión de cobranza”

Michelle Cabrera<sup>1</sup>, James Arias<sup>2</sup>

*Facultad de Ciencias Químicas, Universidad de Cuenca, Ecuador*

---

**Resumen:** La venta a crédito con financiamiento directo es un componente importante en el desarrollo del sector *retail*. A más del riesgo crediticio propio del sector, la presencia de COVID-19 ha traído como consecuencia un incremento en los índices de morosidad debido a un cambio de comportamiento de pago de los clientes por la situación económica del país. Una gestión de cobranza mal definida puede llevar a la toma de estrategias equivocadas y costosas, al desperdicio de recursos y a la desnaturalización de la misma. Esta investigación propone un modelo de segmentación de cartera a partir del comportamiento de los clientes para mejorar la eficiencia de gestión de cobranza. Adoptar este enfoque es clave para generar políticas de gestión y establecer estrategias de cobro específicas. El modelo contempla las variables estáticas más representativas del perfil del cliente provenientes del buró, y atributos dinámicos internos de comportamiento. Se considera un enfoque de clusterización no jerárquica, siendo K-medias la metodología a ser utilizada. El modelo se diseña en Azure Machine Learning y concluye en el software estadístico JMP por el entendimiento gráfico que brinda. Los resultados evidencian una solución distinta al ordenamiento tradicional, ya que comúnmente la clasificación de clientes se realiza en virtud del grado de envejecimiento de cartera. Además, al ser la distancia el discriminante en la formación de *clusters*, ésta puede ser el elemento que facilite el ordenamiento, pues prioriza la pertenencia a uno u otro *cluster*; por lo tanto, es factible usar este concepto como enfoque de prioridad para la gestión de cobranza.

**Palabras clave:** Gestión de cobranza. K-medias. *Retail*. Segmentación

---

<sup>1</sup> Estudiante de Ingeniería Industrial, correo electrónico: [michellecab.cm@gmail.com](mailto:michellecab.cm@gmail.com)

<sup>2</sup> Docente de la Carrera de Ingeniería Industrial, correo electrónico: [james.arias@ucuenca.edu.ec](mailto:james.arias@ucuenca.edu.ec)

**Abstract:** Credit sales with direct financing is an important component in the retail sector development. In addition to the sector's own credit risk, the presence of COVID-19 has resulted in an increase in delinquency rates due to a change in customer payment behavior as a consequence of the country's economic situation. An ill-defined collection management can lead to taking wrong and costly strategies, wasting resources and distorting it. This research proposes a portfolio segmentation model based on customer behavior to improve collection management efficiency. Adopting this approach is key to generating management policies and establishing specific collection strategies. The model contemplates the most representative static variables of the client's profile from the bureau, and internal dynamic behavioral attributes. A non-hierarchical clustering approach is considered, with K-means being the methodology to be used. The model is designed in Azure Machine Learning and concluded in JMP for the graphical understanding it provides. The results show a different solution to the traditional ordering, since the classification of clients is commonly based on the degree of portfolio aging. In addition, since distance is the discriminant in the formation of clusters, this may be the element that facilitates ordering, since it prioritizes belonging to one or another cluster; therefore, it is feasible to use this concept as a priority approach for collection management.

**Keywords:** Collection management. K-means. Retail. Segmentation

## Índice

<b>1. Introducción .....</b>	<b>7</b>
<b>1.1 Segmentación de cartera.....</b>	<b>9</b>
<b>1.2 Clusterización.....</b>	<b>11</b>
1.2.1 Clusterización jerárquica .....	11
1.2.2 Clusterización no jerárquica .....	11
<b>1.3 K-medias .....</b>	<b>12</b>
<b>2. Materiales y Métodos.....</b>	<b>13</b>
2.1 Descripción y procesamiento de los datos .....	14
2.2 Selección de variables .....	15
2.3 Generación del Modelo .....	17
2.4 Clasificación de la cartera .....	18
<b>3. Resultados y Discusión .....</b>	<b>19</b>
<b>3.1 Resultados.....</b>	<b>19</b>
3.1.1 Descripción de los clusters .....	21
3.1.2 Distancia de los clusters al centroide .....	23
<b>3.2 Discusiones .....</b>	<b>23</b>
<b>4. Conclusiones .....</b>	<b>25</b>
<b>5. Bibliografía .....</b>	<b>27</b>

## Cláusula de licencia y autorización para publicación en el Repositorio Institucional

---

Michelle Estefanía Cabrera Calderón en calidad de autor/a y titular de los derechos morales y patrimoniales del trabajo de titulación "Modelo de segmentación de cartera basado en el comportamiento de clientes para la gestión de cobranza", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 21 de julio de 2022.



Michelle Estefanía Cabrera Calderón

C.I: 010463841-6

## Cláusula de Propiedad Intelectual

Michelle Estefanía Cabrera Calderón, autor/a del trabajo de titulación "Modelo de segmentación de cartera basado en el comportamiento de clientes para la gestión de cobranza", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor/a.

Cuenca, 21 de julio de 2022.



---

Michelle Estefanía Cabrera Calderón

C.I: 010463841-6

## 1. Introducción

La estrategia del sector *retail* se basa principalmente en promover el otorgamiento de créditos con un nivel de riesgo superior a la banca y las cooperativas, por medio de mayores facilidades para el acceso a los mismos, en función de modelos de valoración de riesgo; donde la gestión de la cartera de crédito es el pilar fundamental de éxito o de fracaso de las organizaciones y ésta debe ser consistente con la estructura del negocio.

Al igual que otros sectores, el sector *retail* ha experimentado los cambios de la naturaleza del comercio y del comportamiento del consumidor derivado de la pandemia de COVID-19. Kabadayi et al. (citado en Jiang y Stylos, 2021) afirma que las prioridades de los consumidores, los puntos de contacto del servicio *retail* y los procesos comerciales se vieron enormemente afectados, ya que el distanciamiento social impuesto por la mayoría de países condujo a modificar las capacidades de gestión de las organizaciones para que sobrevivieran así a la pandemia.

En una situación posterior al COVID-19 se ocasiona una incertidumbre económica, la cual repercute en la cadena de pagos. Mediante un indicador de Impacto Económico, Equifax (2020) determinó que el 35% de la Población Económicamente

Activa (PEA) del Ecuador, podría ver afectada su liquidez por los efectos económicos de la pandemia, a pesar de haber tenido una situación financiera estable previo al COVID-19. Aún en un contexto de crisis, un 70% de quienes se encuentran en los segmentos menos vulnerables sigue cancelando sus obligaciones; sin embargo, en los segmentos más afectados, el 53% de los ecuatorianos no pagó sus deudas en marzo 2020 (Equifax, 2020).

Análogamente, se identifica que la probabilidad de default por sujeto incrementó en el lapso de Post Pandemia, es decir, el riesgo de los clientes aumentó debido a la situación económica del país: el sobreendeudamiento del mercado crediticio (Equifax, comunicación personal, 2 de febrero de 2022).

Por consiguiente, el riesgo del consumo a crédito propio del sector *retail* y la presencia de COVID-19 en el mundo, tienen como resultado un incremento en los índices de morosidad, obligando de esta forma a redefinir el modelo operativo de cobranza.

Wittlinger, Carranza y Mori (2008) mencionan que la gestión de cobranza es el conjunto de acciones coordinadas y oportunas para lograr la recuperación de los créditos de la manera más rápida y eficiente posible. Durante el proceso de cobranza se

debe decidir cómo proceder con la gestión, considerando la variabilidad de los clientes y de la situación que condujo a la mora. En este sentido, una gestión de cobranza mal definida puede llevar a la toma de estrategias equivocadas y costosas, al desperdicio de recursos y a la desnaturalización de la misma.

Los esfuerzos de las áreas de cobranza cada vez están más enfocados en el manejo preciso de la información relacionada con la medición del desempeño, a fin de fomentar la implementación de acciones concretas (Deloitte, 2012).

Entre las mejores prácticas de gestión de cobranza se recomienda la recopilación y mantenimiento de información, y el establecimiento de políticas, procedimientos y alternativas de pago en base a la segmentación de clientes. Coronel (2019), afirma que, al no segmentar los clientes los resultados de la gestión pueden generar respuestas no favorables o que el costo de la misma supere al beneficio obtenido. Es por esto que una estrategia de cartera segmentada provee herramientas de gran valor al proceso de cobranza.

El criterio de segmentación de clientes junto con la cantidad y calidad de la información que disponga la institución, permiten implementar acciones concretas optimizando los recursos financieros y humanos; para ello es indispensable

considerar aquellas variables que mejor expliquen la actuación del cliente. Así mismo, es importante el seguimiento de las características y comportamiento del mismo empleando técnicas estadísticas que faciliten su análisis; y el control de los procesos de recuperación de cartera que ajuste e impida la desviación de las políticas establecidas y que incorpore un sistema de monitoreo para la consecución de los objetivos. El problema de seleccionar segmentos diversificados correctamente se puede abordar mediante la clusterización (Nanda, Mahanty y Tiwari, 2010).

Los algoritmos de clusterización están siendo utilizados en el mundo empresarial y en la comunidad científica, facilitando el análisis estadístico y haciéndolo más eficiente. La clusterización es una técnica de aprendizaje no supervisado que revela estructuras de grupos de datos. Las variables que pertenecen a un grupo están más relacionadas que las de otros grupos, de modo que las observaciones ya no se consideran únicas, sino que se ven como miembros de grupos que pueden describirse mediante perfiles característicos.

Con el fin de mejorar la eficiencia de la gestión de cobranza, en este artículo se utiliza clusterización para segmentar la cartera de clientes. La técnica de clusterización utilizada es K-medias. En base a índices estadísticos de clusterización



y criterios de selección se define el número óptimo de *clusters*. La base de datos contiene atributos de comportamiento, así como de características crediticias. Por lo tanto, los grupos resultantes son una clasificación de datos multivariantes de gran dimensión en diferentes grupos en vista de la diferencia entre las variables más representativas.

El documento está organizado de la siguiente manera. El resto de la sección 1 describe la revisión de la literatura relevante. La sección 2 presenta la metodología utilizada. En la sección 3 se describen los resultados obtenidos. Finalmente, la sección 4 presenta la conclusión.

## 1.1 Segmentación de cartera

La industria financiera está recurriendo a nuevas técnicas y herramientas de análisis para alcanzar una evaluación más precisa del riesgo asociado a cada cliente. Los enfoques de segmentación utilizados han cubierto una amplia gama de metodologías que consideran las variables del comportamiento histórico del cliente, con las que se busca predecir su conducta de pago. Según Deloitte (2012), *Behavioural* y *Collection Scoring* están entre los modelos más utilizados para definir segmentos y clasificar a los clientes.

Entre los beneficios del modelo de segmentación de cobranza se encuentran: el incremento del porcentaje de recuperación

de cartera vencida, el direccionamiento de esfuerzos para la cartera con mayor probabilidad de recuperación, la creación de estrategias para los distintos segmentos de clientes, y la mejora en la eficiencia de los procesos de cobranza; lo cual orientará a una labor de administración y análisis del desempeño (Deloitte, 2012).

*Behavioural scoring* es una calificación en donde los datos utilizados para el ajuste del modelo se basan en el desempeño del reembolso del préstamo (Kennedy, Mac Namee, Delany, O'Sullivan y Watson, 2013).

El uso de *scoring* en entidades financieras es una técnica muy utilizada principalmente en la admisión del crédito. Sin embargo, en la gestión de cobranzas se emplea para pronosticar eventos en base a probabilidades, otorgando puntajes a variables conocidas; de esta manera, se mantiene un sistema preventivo de teoría de colas generando tramos de cobranzas para determinar estrategias en función del perfil de los clientes (Riera, 2018).

Para establecer estrategias focalizadas que incrementen los niveles de recuperación de la cartera de microcréditos y optimicen los recursos, Rodríguez (2018) plantea un modelo de *scoring* a través de la discriminación entre clientes buenos y malos mediante cuatro algoritmos de aprendizaje

supervisado, concluyendo que el mejor modelo es por medio de árboles de decisión con un error del 11,3%. Las variables utilizadas se dividen en dos grupos: perfil del cliente (edad, monto, antigüedad del crédito, nivel de ingresos, total de activos, garantía, género y sector urbano, rural o agropecuario), y comportamiento histórico de los últimos 6 meses (porcentaje de cuotas pagadas, máxima mora, número de veces de mora, pagos realizados).

Riera (2018) desarrolla un modelo de *scoring* de segmentación de cobranzas para tarjetas de crédito utilizando la regresión logística binaria *Logit*. La variable dependiente es el incumplimiento en función de los días mora (límite de mora de 61 días), siendo ésta una variable dicotómica (cliente bueno o malo), la misma es identificada previo a la construcción del modelo. Al utilizar una regresión *Logit* se tiene como ventaja la determinación probabilística de este tipo de comportamiento. Es decir, el modelo no proporciona directamente los segmentos, sino la probabilidad de mora de los clientes, con la cual se establecen puntos de corte para realizar la segmentación. Las variables identificadas como influyentes son características propias del comportamiento de los clientes (información interna) con las cuales se crean nuevas variables de agrupación potencialmente explicativas, y datos del buró de crédito (información

externa); entre ellas están: incidencias previas de mora, días de mora promedio, calificación buró, *score* interno, saldo deuda vs cupo, y cantidad de avances. El modelo presenta un porcentaje de error inferior al 5%, con los resultados obtenidos se elabora una matriz ordenada de cobranzas basada en los segmentos, considerando como referencia las variables incluidas en el modelo.

Los modelos pueden utilizar características estáticas sobre el desempeño pasado del cliente y/o técnicas que incorporan aspectos dinámicos. Kennedy et al. (2013) utilizan características estáticas de comportamiento centrado en datos provenientes del buró al momento de la aplicación del crédito, e incluyen indicadores de pago para su modelo de discriminación de regresión logística. Para construir el modelo, la variable dependiente (cliente bueno o malo) está definida a partir del límite de días de mora establecido. La investigación sugiere un enfoque adicional de definición de variable dependiente basado en un porcentaje del monto de los atrasos y el valor pendiente del préstamo.

Thomas, Ho y Scherer (citado en Kennedy et al., 2013) analizan los enfoques de la calificación conductual, con una perspectiva particular en los procedimientos que incorporan aspectos dinámicos del comportamiento del cliente, especialmente

el estado de la morosidad en distintos periodos.

Las variables identificadas como influyentes en los modelos de segmentación de cartera, tienen una relación directa y significativa con el perfil del cliente y con el comportamiento histórico de pagos del mismo. En las investigaciones existentes, los modelos desarrollados son de aprendizaje supervisado, es decir, determinan la probabilidad de un evento a partir de datos de entrenamiento, en donde la variable dependiente está etiquetada previamente y en base a esta etiqueta se entrena el modelo que se utilizará con el nuevo conjunto de datos; por lo general, estos modelos determinan la probabilidad de pago del cliente y en función de ésta se definen los segmentos a criterio de las organizaciones. Sin embargo, para organizaciones del sector *retail* no se observan modelos de gestión de cartera segmentada.

## 1.2 Clusterización

El objetivo de los algoritmos de clusterización es colocar los datos en grupos en función de las variables seleccionadas, en los que los individuos de un grupo sean más similares entre sí que a los de otros grupos. La clusterización es un algoritmo de aprendizaje no supervisado, donde se trabaja con un conjunto de datos que no tienen etiquetas asociadas a ellos. Los métodos de

clusterización se caracterizan por diferentes formas de medir la homogeneidad y diversos procedimientos para encontrar la partición óptima.

### 1.2.1 Clusterización jerárquica

En los algoritmos de clusterización jerárquica, cada una de las  $n$  observaciones se considera un *cluster* único. Posteriormente, se comparan entre sí y se asocian los grupos con menor distancia. Este procedimiento es repetitivo hasta que se alcanza el número de *clusters* deseado (Mingoti y Lima, 2006).

### 1.2.2 Clusterización no jerárquica

A diferencia de los algoritmos jerárquicos, el número deseado de *clusters*  $k$  tiene que estar predefinido. El propósito es agrupar las  $n$  observaciones en  $k$  *clusters* de tal manera que los datos del mismo grupo sean similares y los que pertenecen a diferentes grupos sean heterogéneos, en las  $p$  variables utilizadas para el algoritmo.

Chiu, Chen, Kuo y He (2009) emplearon K-medias para la segmentación de mercado en la toma de decisiones de estrategia. Nanda et al. (2010), sugieren cómo integrar técnicas de agrupación como K-medias, SOM y Fuzzy C-medias en la gestión de cartera; posterior al análisis de los métodos, determinan que K-medias resulta ser el enfoque más robusto en base a la solidez de los *clusters*, siendo esto similar a

los resultados obtenidos por Shin y Sohn (2004) para la segmentación de clientes. Análogamente, Mingoti y Lima (2006) **consideran a K-medias como el método más adecuado para agrupar datos multivariantes**, ya que en comparación con SOM y Fuzzy C-medias, se obtuvo grupos más lógicos y mutuamente excluyentes, a diferencia de los otros métodos analizados.

### 1.3 K-medias

El método de clusterización K-medias es uno de los métodos más utilizados en el procesamiento de datos a gran escala debido a su simplicidad, rapidez y eficiencia (Zhang, Yang et al. citado en Yang, Long, Ma y Sun, 2020). El algoritmo empieza con la selección de  $k$  puntos multidimensionales para que sean las “semillas” o centroides iniciales para los  $k$  clusters que se formarán. Luego, se compara la distancia de las  $m$  observaciones con cada centroide y se asigna cada observación al grupo con el centroide más cercano. El proceso es iterativo y el centroide de cada cluster se vuelve a calcular utilizando el vector promedio de las observaciones asignadas al cluster. El algoritmo se detiene cuando los cambios en los centroides del grupo sean pequeños.

Entre las métricas utilizadas para calcular las distancias entre las observaciones y los centroides, están:

*Distancia Manhattan*: es una métrica que calcula las diferencias absolutas entre coordenadas (Singh, Yadav y Rana, 2013).

*Distancia euclidiana*: calcula la raíz cuadrada del cuadrado de las diferencias entre coordenadas; la misma es comúnmente usada como una medida de dispersión de grupos para el agrupamiento de K-medias, pues esta métrica minimiza la distancia media entre los puntos y los centroides (Microsoft, 2006).

*Distancia Minkowski*: es una métrica en un espacio vectorial normalizado que puede considerarse como una generalización tanto de la distancia euclidiana como de la distancia de Manhattan (Singh et al., 2013).

*Distancia Mahalanobis*: se diferencia de la distancia euclídea en que tiene en cuenta la correlación entre las variables (Escobedo y Salas, 2008).

La medición del resultado de la agrupación se realiza a través de índices de validez que determinan el número óptimo de clusters mutuamente excluyentes entre sí, y menos dispersos dentro de cada uno de ellos. Algunos de los índices de validación se discuten brevemente.

*Índice de silueta (Simplified Silhouette)*: captura la rigidez de los puntos de datos dentro de cada cluster. La mejor calidad de clusterización se indica mediante

un valor de silueta más grande (Chen et al. citado en Nanda et al., 2010).

*Índice de Davies-Bouldin:* identifica el conjunto más pequeño de *clusters* con la menor dispersión, por lo tanto, un valor más bajo indica un mejor agrupamiento (Chen et al. citado en Nanda et al., 2010).

*Índice de Dunn:* identifica el conjunto más pequeño de los *clusters* más compactos. Un valor más alto para esta métrica indica la presencia de *clusters* compactos y bien separados (Bezdek y Pal citado en Nanda et al., 2010).

## 2. Materiales y Métodos

La mayoría de cadenas comerciales que otorgan crédito directo, realizan una clasificación tradicional de clientes únicamente en función del número de días de mora; por ello, es a lo que el estudio pretende dar una segunda opción, ya que a partir de un cúmulo de variables se aspira proveer de gran valor al proceso de cobranza.

Como caso de estudio se busca agrupar los datos y dar un ordenamiento específico con objeto de mejorar la eficiencia de gestión de cobranza asociada a la segmentación de cartera a partir del conocimiento previo de las características de los clientes.

Para la evaluación de lo propuesto, la investigación se enfoca en una empresa

*retail* dedicada a la comercialización de motocicletas y electrodomésticos a nivel nacional, en la cual la venta a crédito representa el 80% de las ventas totales de la empresa. La misma posee una cartera no segmentada con más de 20.000 clientes, distribuidos en 30 ciudades del Ecuador.

Actualmente, la asignación de la cartera a gestionar se lo realiza de manera mensual y está distribuida en 12 tramos. La asignación de éstos se realiza en base al número de días de atraso, con respecto a la fecha que debió haber sido cancelada la cuota más antigua. El tramo puntual está conformado por todos los créditos recién otorgados y por aquellos que no registran atraso en la cuota correspondiente, mientras que la cartera vencida está constituida por todos los clientes que incurren en mora. El no cancelar el valor de una cuota completa en el mes correspondiente, ocasiona un desplazamiento de tramo para el próximo mes.

Para las gestiones de cartera, ésta se divide en dos grupos: cartera puntual y cartera vencida. La distribución de cartera a gestionar se realiza en función de la política de prioridades determinada por dos variables: tipo de cliente y agrupación por fecha de venta. El esquema de prioridades presentado en la Figura 1, está vigente para cada uno de los 12 tramos existentes, por lo tanto, el número de grupos de atención es amplio, siendo 48 el total.

TIPO DE CLIENTE	GRUPO FECHA VENTA	
	SIEMBRA	NO SIEMBRA
CLEAN	Prioridad 1	Prioridad 2
DIRTY	Prioridad 3	Prioridad 4

Figura 1. Esquema de prioridades para la gestión de cobranza

Tipo de cliente es una variable (dato estático externo) obtenida del buró de crédito que indica el comportamiento “limpio” o “sucio” del cliente, respecto a créditos anteriores. El comportamiento crediticio “Clean” está constituido por quienes, en los últimos 36 meses, no han presentado un vencimiento mayor a 30 días, a diferencia del comportamiento “Dirty”.

Mientras que, la agrupación por fecha de venta es una variable (dato estático interno) que indica si los créditos fueron otorgados pre (siembra) o post (no siembra) pandemia de COVID-19.

La gestión de cobranza se realiza a través de recursos internos: almacenes, recaudadores, *contact center* y departamento legal; y recursos externos: abogados extrajudiciales y empresas recaudadoras. Hoy en día, se ha identificado multiplicidad de gestiones de diversos orígenes para un mismo grupo de clientes. Adicional, no todas las gestiones proporcionan los mismos resultados en todos los clientes.

En consecuencia, en la estructura

actual para el proceso de cobranza, únicamente se utilizan datos estáticos para establecer las prioridades de gestión; independientemente del tramo, el mismo que es un dato dinámico pues varía con cada corte de cartera. De ahí que, el total de grupos para la gestión es 48.

## 2.1 Descripción y procesamiento de los datos

La base de datos proporcionada contiene datos estructurados pertenecientes a los cortes mensuales de cartera que corresponden al período agosto 2020 - junio 2021, debido a la relevante variación de comportamiento de pago post COVID-19. La misma, está constituida por 338.516 registros de comprobantes distribuidos en los distintos tramos, con información externa relacionada al perfil crediticio del cliente, e información interna asociada a su comportamiento histórico con respecto al crédito otorgado. Las variables contempladas en la base de datos se muestran en la Tabla 1.

Adicional, las variables presentes en la base tienen dos consideraciones: variables estáticas y variables dinámicas. Las variables estáticas son atributos descriptivos del crédito, y se constituyen al momento en el que se otorga el mismo, así pues, éstas no varían con cada corte.



**Tabla 1**

*Variables independientes*

TIPO DE VARIABLE	ESTÁTICA	DINÁMICA	
INTERNA	Código de comprobante	Cartera castigada	
	Tipo de comprobante	Tramo	
	Código de persona	Fecha último pago	
	Nombre de persona	Cuotas activas	
	Fecha de venta	Código de mes	
	Política	Descripción de mes	
	Día de vencimiento	Saldo inicial	
	Plazo	Interés de mora inicial	
	Primer vencimiento	Saldo	Asignación
			Saldo cancelado
			Fecha pago mes
EXTERNA	Tipo de cliente		
	Segmentación		
	Score		

Por el contrario, las variables dinámicas son calculadas al corte de cartera, esto es, de manera mensual. En consecuencia, cada corte ocasiona un cambio en el comportamiento de los clientes; en otros términos, un mismo cliente puede tener  $n$  observaciones, ya que cada corte genera nuevos datos.

## 2.2 Selección de variables

Las variables identificadas como influyentes en los modelos de segmentación de cartera, tienen una relación directa con el perfil del cliente y con el comportamiento histórico de pagos del mismo; es por esta razón que para la obtención de las variables e índices que garanticen una segmentación que se ajuste a la composición de la cartera, se utilizan las variables externas provenientes del buró

de crédito, y se establecen relaciones entre los datos internos existentes, con los cuales se crean nuevas variables de comportamiento potencialmente explicativas.

De las 21 variables internas iniciales, de la Tabla 1, se excluyen del modelo los 8 atributos estáticos y 9 dinámicos, presentados en la Tabla 2, debido a que los mismos representan información del crédito otorgado, de manera que son utilizados para generar conteos, determinar ordenamientos, establecer temporalidades y diferenciar las distintas observaciones presentes en la base de datos.

**Tabla 2**

*Atributos internos descartados*

TIPO DE VARIABLE	ESTÁTICA	DINÁMICA	
INTERNA	Código de comprobante	Cartera castigada	
	Tipo de comprobante	Tramo	
	Código de persona	Fecha último pago	
	Nombre de persona	Cuotas activas	
	Fecha de venta	Código de mes	
	Política	Descripción de mes	
	Plazo	Interés de mora inicial	
	Primer vencimiento	Saldo	Asignación

Posterior a esta segregación inicial, se procede a trabajar con las variables de la Tabla 3, las mismas se dividen en dos grupos: variables de perfil del cliente (datos externos) y de su comportamiento histórico de pago (datos internos).

**Tabla 3**

*Variables disponibles*

TIPO DE VARIABLE	ESTÁTICA	DINÁMICA
INTERNA	Día de vencimiento	Saldo inicial
		Saldo cancelado
		Fecha pago mes
EXTERNA	Tipo de cliente	
	Segmentación	
	Score	

En la Tabla 4, se presentan las variables de perfil del cliente a usarse en el modelo, las mismas hacen referencia a la información obtenida del buró, ingresada al momento de la aprobación del crédito; es decir, son datos estáticos externos.

**Tabla 4**

*Variables de perfil del cliente*

TIPO DE VARIABLE	ESTÁTICA
EXTERNA	Tipo de cliente
	Segmentación
	Score

Tipo de cliente indica la presencia o ausencia de un vencimiento mayor a 30 días en los últimos 36 meses, en el sistema crediticio; la Segmentación es una clasificación basada en la probabilidad de incumplimiento de pago, y el Score equivale al puntaje de riesgo del cliente.

En contraste, las variables de comportamiento de pago brindan información histórica del desempeño de

pago del cliente frente a las obligaciones de crédito contraídas con la empresa, es decir, son datos internos.

A partir de las variables internas disponibles, se crean nuevas variables dinámicas de agrupación que conducen a un mejor rendimiento del algoritmo. Las mismas se muestran en la Tabla 5.

**Tabla 5**

*Variables de comportamiento de pago*

TIPO DE VARIABLE	DINÁMICA
INTERNA	Recuperación en saldo
	Activación
	Índice de pago

La recuperación en saldo representa la relación entre el saldo cancelado y saldo inicial asignado en el mes de corte, tal como se indica en la Ecuación 1. La finalidad de este atributo es exponer el cumplimiento de las obligaciones para cada mes de corte.

$$R_{\text{saldo}} = \frac{\text{Saldo cancelado}}{\text{Saldo inicial}} \quad (1)$$

Como se observa en la Ecuación 2, la activación relaciona el número de asignaciones y el número de cancelaciones para un mismo crédito, con la intención de determinar el compromiso frente a las obligaciones contraídas.

$$\text{Activación} = \frac{\text{Cancelaciones}}{\text{Asignaciones}} \quad (2)$$



Finalmente, el índice de pago es la diferencia entre el grupo de pago y el grupo de vencimiento, así como se manifiesta en la Ecuación 3. Tras esta variable se identifica la puntualidad de las cancelaciones en el mes de corte.

$$\text{Índice pago} = G.Pago - G.Venc \quad (3)$$

En definitiva, a partir de la formulación y selección de variables del conjunto inicial conformado por 24 atributos, se determina que 6 son las más representativas para aplicar en el algoritmo. Los atributos descartados son datos estáticos internos que no concuerdan con el objetivo de incorporar aspectos dinámicos de desempeño en el algoritmo; mientras que las demás variables que se excluyen son utilizadas para el cálculo de variables que en efecto están siendo consideradas el algoritmo.

En la Tabla 6, se observan las variables a emplearse en el modelo.

**Tabla 6**

*Variables para el modelo*

TIPO DE VARIABLE	ESTÁTICA	DINÁMICA
INTERNA		Recuperación en saldo
		Activación
		Índice de pago
EXTERNA	Tipo de cliente	
	Segmentación	
	Score	

## 2.3 Generación del Modelo

El modelo que se contempla en esta investigación pretende solventar el problema de la segmentación para la gestión de cobranza, es por ello que se realiza una clusterización con la cual se aspira segmentar la cartera en *clusters* generados a partir de valores predominantes de las variables empleadas en el algoritmo. Dicho análisis se realiza a través de un modelo de aprendizaje automatizado, concretamente con clusterización no jerárquica acorde a lo sugerido en los estudios detallados en la sección 1.

El modelo se diseña en Microsoft Azure Machine Learning. Esta plataforma es una herramienta de bajo código por lo que simplifica y acelera el proceso de creación, prueba e implementación de modelos de aprendizaje automatizado. Y termina en JMP por el entendimiento gráfico que brinda este software estadístico.

Adicional, se evalúa la distancia de las observaciones a los *clusters*, con objetivo de obtener criterios de ordenación que indican la prioridad para la gestión de cobranza.

En primer lugar, en función de la variable “tramo” se establecen dos comportamientos que permiten la división de la base de datos en dos grupos: comportamiento puntual y comportamiento

*default*, debido a que existe una divergencia en la conducta de los clientes que se mantienen siempre puntuales de aquellos que incurren en mora. Esta división es fundamental, ya que al no realizarla tergiversa la formación de *clusters*. Por lo tanto, el modelo se aplica para cada uno de los grupos.

En esta investigación, se considera un enfoque de clusterización no jerárquica, siendo K-medias la metodología a ser utilizada para la segmentación de clientes; tomando en consideración la distancia euclidiana como medida para calcular la distancia entre cada uno de los puntos del conjunto de datos a los centroides. Se emplea la distancia euclídea pues es comúnmente usada como una medida de dispersión de grupos para el agrupamiento de K-medias. Además, se prefiere esta métrica porque minimiza la distancia media entre los puntos y los centroides, tal como se indica en la sección 1.3.

El algoritmo empieza con la selección de  $k$  puntos multidimensionales, para que sean los centroides iniciales para los  $k$  *clusters* que se formarán. El número óptimo de *clusters* se obtiene mediante el análisis de índices de validez y criterios de selección.

Posteriormente, se procede a ejecutar el modelo de clusterización que se muestra en la Figura 2 para el comportamiento

puntual, y en la Figura 3 para el comportamiento *default*.

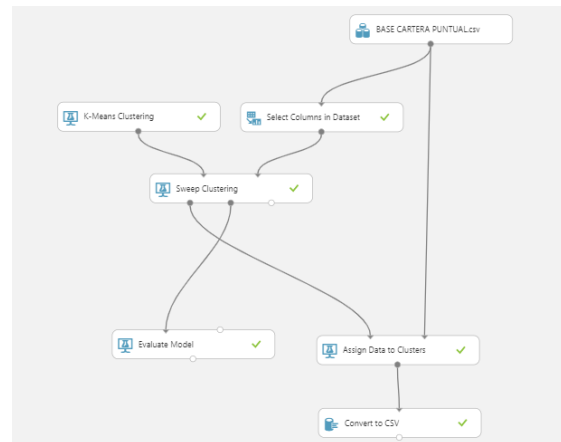


Figura 2. Modelo de clusterización para el comportamiento puntual

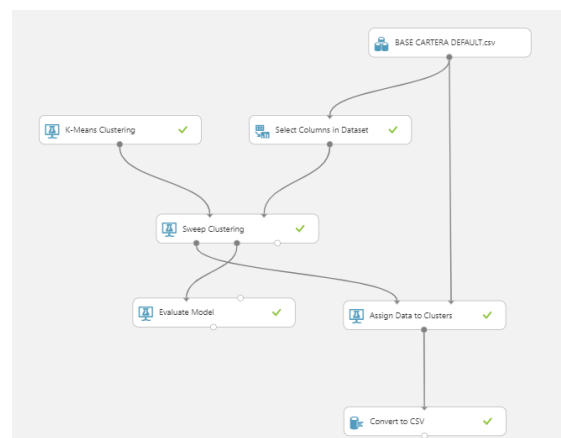


Figura 3. Modelo de clusterización para el comportamiento *default*

## 2.4 Clasificación de la cartera

A inicio de mes, la totalidad de la cartera asignada se clasifica mediante el modelo entrenado, a través de la extensión de publicación de Microsoft Azure Machine Learning, en Microsoft Excel; y se liga la misma a las políticas y estrategias de

cobranza de la empresa, en función de los criterios de ordenación establecidos por la organización.

### 3. Resultados y Discusión

#### 3.1 Resultados

A partir de los datos obtenidos, se identificó dos comportamientos: puntual y *default*. El primer comportamiento parte de la situación en la que los clientes se han mantenido siempre puntuales en el pago de sus cuotas; por el contrario, el comportamiento *default* se caracteriza por una tendencia a la morosidad, es decir, haber incurrido en mora al menos una vez.

Esta discriminación inicial es primordial para no distorsionar la formación de *clusters*. Por lo tanto, los resultados obtenidos se manifiestan para cada uno de los grupos.

A continuación, con las variables más representativas del perfil del cliente y de su

comportamiento histórico de pago, especificadas en la sección 2.2, se determinó el número óptimo de *clusters* mediante el análisis de índices de validez como se indica en la sección 1.3.

El número de *clusters* a evaluarse, tuvo un rango de variación de 3 a 8, para ambos comportamientos. Los índices de validez interna se calcularon y tabularon en las Tablas 7 y 8 para el comportamiento puntual y *default*, respectivamente.

Mediante el análisis de la Tabla 7, se dedujo que para el grupo de comportamiento puntual el número óptimo de *clusters* es cuatro, pues cumple con las condiciones de selección en todos los índices: el Índice de Silueta y el Índice de Dunn son más altos, por lo que reflejan una mejor calidad de clusterización, y *clusters* compactos y mutuamente excluyentes, respectivamente; y el Índice de Davies-Bouldin es más bajo, representando el conjunto de *clusters* con la menor dispersión.

**Tabla 7**

*Índices de validez del agrupamiento de K-medias para el comportamiento puntual*

Índices de validez	Número de clusters					
	3	4	5	6	7	8
Comportamiento puntual						
Índice de Silueta	0.5573	0.5700	0.5506	0.5258	0.5064	0.5031
Índice de Davies-Bouldin	0.8702	0.8516	1.0979	1.1968	1.2672	1.3192
Índice de Dunn	0.6962	0.8108	0.7727	0.6881	0.6137	0.4873

**Tabla 8**

*Índices de validez del agrupamiento de K-medias para el comportamiento default*

Índices de validez	Número de clusters					
	3	4	5	6	7	8
Comportamiento <i>default</i>						
Índice de Silueta	0.5607	0.6028	0.6092	0.5469	0.5321	0.5884
Índice de Davies-Bouldin	0.9047	0.8356	0.8114	1.0224	1.0292	0.9282
Índice de Dunn	0.5150	0.4929	0.5795	0.2884	0.2811	0.3074

Por su parte, para el grupo de comportamiento *default*, en la Tabla 8 se evidencia que el número de *clusters* para un agrupamiento óptimo de los datos es cinco, dado a que el Índice de Silueta y el Índice de Dunn son más altos, y el Índice de Davies-Bouldin es más bajo, respetando así las condiciones de los índices de validez que determinan el número óptimo de *clusters* mutuamente excluyentes entre sí, y menos dispersos dentro de cada uno de ellos.

Finalmente, se procedió a ejecutar el modelo de clusterización propuesto con la cantidad de *clusters* establecidos para cada grupo de comportamiento.

Los resultados para el comportamiento puntual se visualizan en las Figuras 4 y 5, donde se precisan claramente los cuatro *clusters* formados, pues la gráfica Biplot muestra la separación y la superposición de los diferentes *clusters*.

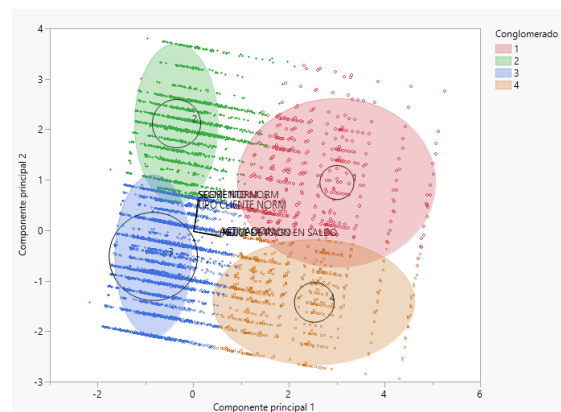


Figura 4. Biplot comportamiento puntual

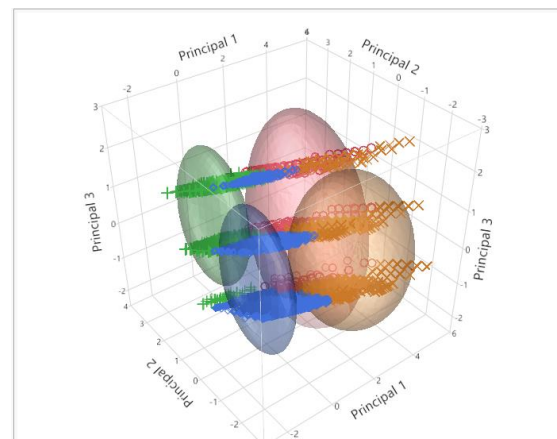


Figura 5. Biplot 3D comportamiento puntual

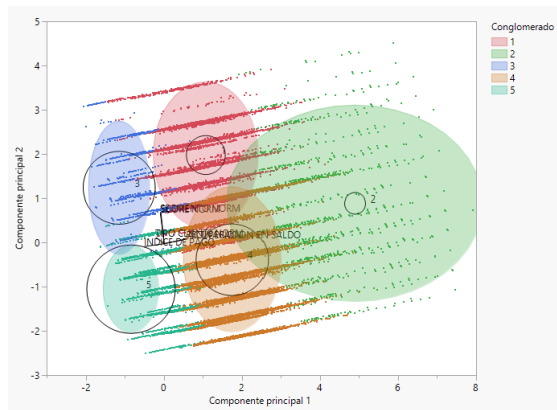
La composición de los clusters formados con respecto al comportamiento puntual, se detalla en la Tabla 9.

**Tabla 9**

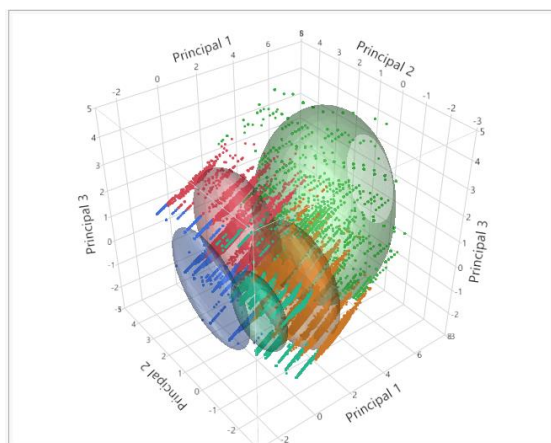
*Resumen de clusterización para el comportamiento puntual*

CLUSTER	TOTAL	% TOTAL
1	7.156	8,32%
2	21.521	25,03%
3	45.341	52,74%
4	11.960	13,91%

En las Figuras 6 y 7, se distinguen los cinco *clusters* obtenidos para el comportamiento *default*, mediante la aplicación del modelo.



*Figura 6. Biplot comportamiento default*



*Figura 7. Biplot 3D comportamiento default*

El resumen de composición de los *clusters* para este comportamiento se detalla en la Tabla 10.

**Tabla 10**

*Resumen de clusterización para el comportamiento default*

CLUSTER	TOTAL	% TOTAL
1	22.858	9,05%
2	9.587	3,80%
3	67.300	26,65%
4	52.141	20,65%
5	100.652	39,86%

En función de los resultados obtenidos, se establecieron las características de los *clusters* generados según los valores predominantes de las variables empleadas en la clusterización. Adicional, se evaluó la distancia de las observaciones a los *clusters*, con objetivo de obtener criterios de ordenación que indican la prioridad para la gestión de cobranza.

### 3.1.1 Descripción de los clusters

En lo que concierne al comportamiento puntual, la Figura 8 representa la ponderación de variables para cada *cluster*. En la misma, se evidencia la diversidad en la composición de los *clusters* formados, tanto en las variables estáticas

externas (perfil de cliente) como en las dinámicas internas (comportamiento).

Las variaciones que destacan en los *clusters* constituidos son:

*Cluster 1:* Calificación crediticia muy baja con tendencia al atraso, por lo tanto, la probabilidad de *default* es mayor.

*Cluster 2:* Calificación crediticia media baja cuyo comportamiento es cancelar en la fecha correspondiente.

*Cluster 3:* Calificación crediticia muy alta con tendencia nula a la mora.

*Cluster 4:* Calificación crediticia media alta con tendencia nula a la mora, pues siempre cancela anticipadamente.

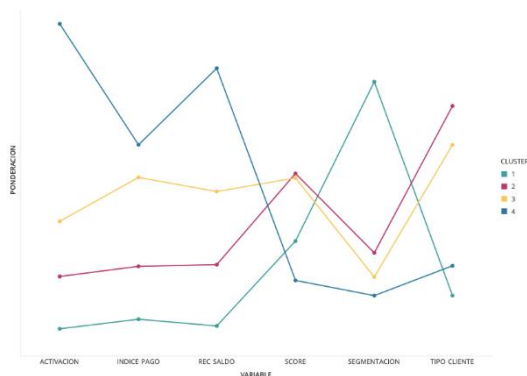


Figura 8. Composición de *clusters*: comportamiento puntual

Análogamente, en la Figura 9 se observa la ponderación de variables con respecto a cada *cluster* perteneciente al comportamiento *default*. La misma destaca la diversificación de los atributos de perfil de cliente y de comportamiento de pago, dentro de los *clusters* formados.

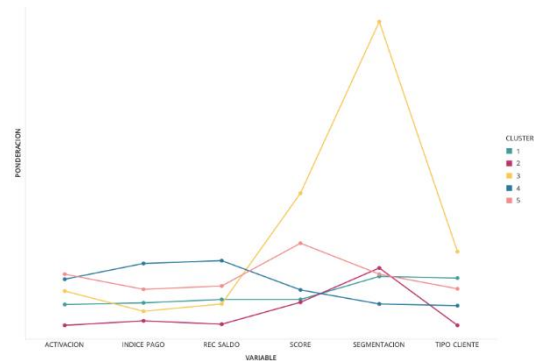


Figura 9. Composición de *clusters*: comportamiento *default*

Dentro del comportamiento *default*, la definición de los *clusters* es la siguiente:

*Cluster 1:* Calificación crediticia baja, cuyo comportamiento es cancelar antes o en la fecha correspondiente.

*Cluster 2:* Calificación crediticia muy baja con tendencia al atraso y activación parcial.

*Cluster 3:* Calificación crediticia media con tendencia al atraso y activación nula.

*Cluster 4:* Calificación crediticia alta cuyo comportamiento general es cancelar en la fecha correspondiente.

*Cluster 5:* Calificación crediticia media alta con activación nula.

La Figura 10 expone la distribución de la antigüedad de la cartera, representada por la variable “tramo”, dentro de los *clusters* constituidos para el comportamiento *default*. En la misma, se evidencia la multiplicidad de tramos presentes en un mismo *cluster*.



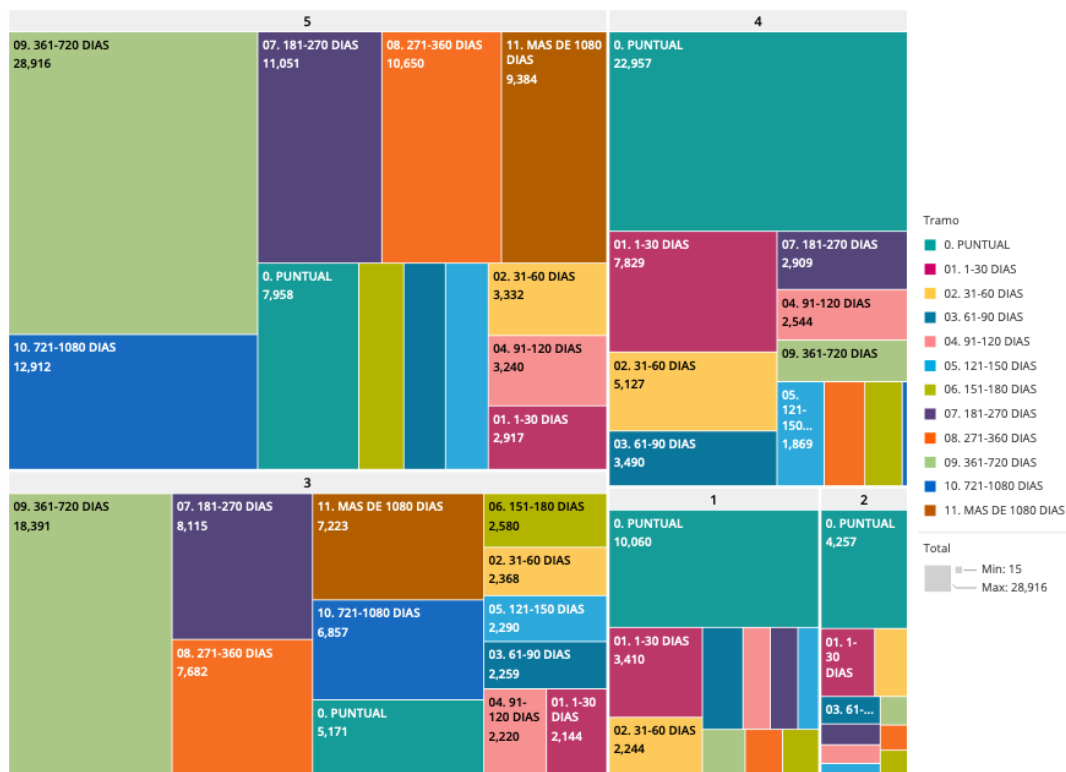


Figura 10. Distribución de tramos: comportamiento *default*

### 3.1.2 Distancia de los clusters al centroide

Los resultados de la clusterización que se obtienen al aplicar el modelo son: el *cluster* al que pertenece la observación en concreto, y la distancia de cada observación a los centroides de cada uno de los *clusters* formados. Entendiendo de esta forma, que la distancia es el factor discriminador en la formación de *clusters*, pues un punto se descarta o no de uno de los conglomerados por la distancia que tenga el mismo con respecto a los *clusters* constituidos.

### 3.2 Discusiones

El uso de un modelo matemático permite a las organizaciones del sector *retail*

tener una mejora en la eficiencia de los procesos de cobranza que dispone pues, según los resultados obtenidos, el diseñar un algoritmo que abarque tanto el perfil crediticio del cliente como el comportamiento histórico de pagos admite una nueva manera de segmentación y ordenamiento específico de la cartera.

La tendencia de cobranza y recuperación de cartera en el sector financiero es la clasificación tradicional de clientes en función de datos estáticos, y/o en virtud del número de días de mora (tramos); de ahí que el estudio proporciona una segunda opción pues, a partir de un cúmulo de variables estáticas y dinámicas, tanto

internas como externas, se provee de gran valor al proceso de cobranza. Originalmente, datos que pertenecen a un tramo en específico se distribuyen en varios *clusters*, puesto que podrían existir clientes considerablemente atrasados, pero pertenecer a un *cluster* de mayor probabilidad de pago, a causa de su comportamiento; evidenciando de esta manera que el modelo propuesto es una solución diferente al ordenamiento tradicional, debido a que anteriormente la utilización de grupos estaba proporcionada por un único factor que es estrictamente el grado de envejecimiento de la cartera. Adicional, la agrupación tradicional divide la cartera en más grupos de atención, mientras que la clusterización realizada en esta investigación en menos conglomerados; por consiguiente, el desafío es cambiar las acciones en función de los esquemas de segmentación establecidos.

Se enfatiza la relevancia de la discriminación inicial acerca de los comportamientos puntual y *default*, dado que los resultados obtenidos para cada uno de ellos difieren significativamente entre sí como se observan en las figuras de composición de los *clusters*.

Referente a los *clusters* formados, se destaca la diversidad presente en los mismos debido a las variables identificadas como influyentes, las mismas que tienen una

relación directa con el perfil del cliente y con el comportamiento histórico de pagos del mismo. Cabe resaltar la importancia de considerar los dos tipos de variables en el estudio, pues en caso de no hacerlo la formación de *clusters* se tergiversaría, ya que existen segmentos similares en cuanto al perfil crediticio pero cuya diferencia radica en el comportamiento histórico que hayan presentado, o viceversa.

Desde un punto de vista lógico, la clusterización equivale a un concepto de grado de atracción por cercanía, debido a que una observación podría no ser de un *cluster* si está lo suficientemente lejos del centroide de éste, y ciertamente sería capturado en otro *cluster* más cercano. Por lo tanto, el grado de cercanía tiene relación directa con la potencia de pertenencia al grupo; dicho en otras palabras, la distancia tiene bastante sentido con respecto a la fuerza de atracción o a la pertinencia de un *cluster*. Bajo esta figura, la distancia puede ser un elemento que facilite el ordenamiento por atracción, el cual a su vez prioriza la pertenencia. Entonces, es factible usar este concepto como enfoque de prioridad para la gestión de cobranza, en función de la prelación asignada a los *clusters*.

Esta investigación aborda el problema de seleccionar segmentos diversificados correctamente, y a su vez plantea un ordenamiento específico para reducir



esfuerzos de las áreas de cobranza y así mejorar la eficiencia de su gestión. A partir de las 338.516 observaciones presentes en la base de datos proporcionada, la misma que equivale a una cartera no segmentada con más de 20.000 clientes, se genera el modelo que combina un matiz de las variables utilizadas en distintos estudios como datos provenientes del buró y datos históricos de pagos; cabe destacar que en varias investigaciones previas no mencionan la utilización en conjunto de atributos estáticos y dinámicos.

En las investigaciones existentes, los modelos desarrollados son de aprendizaje supervisado, es decir, determinan la probabilidad de un evento a partir de datos de entrenamiento, en donde la variable dependiente está definida con anterioridad y en base a ésta se entrena el modelo, que, por lo general, establecen la probabilidad de pago del cliente y en función de ésta se definen los segmentos a criterio propio. Por el contrario, en el modelo propuesto se emplea aprendizaje no supervisado, pues a partir de un conjunto de observaciones no etiquetadas previamente, se revela estructuras de grupos de datos. Simultáneamente, se busca llenar un vacío existente pues no se observan modelos de gestión de cartera segmentada para organizaciones del sector *retail*.

## 4. Conclusiones

Al ser la segmentación de cartera un factor crítico en la gestión de cobranza, implementar un modelo que utiliza el perfil crediticio y el historial de comportamiento interno del cliente, se convierte en una forma de optimizar y orientar acciones concretas en el proceso de cobranza.

La utilización de variables estáticas externas provenientes del buró crediticio, y atributos dinámicos internos de comportamiento, ayuda a determinar relaciones entre observaciones que no son inmediatamente notorias cuando se consideran las variables individualmente; además, favorecen la predicción de la probabilidad de pertenencia de un individuo concreto a uno de los segmentos diferenciados.

A diferencia de la clasificación tradicional de cartera, que es únicamente en función de datos estáticos o conforme al tramo definido por el número de días de mora, con varios grupos de atención resultantes; la clusterización realizada en esta investigación, da origen a la formación de una cantidad menor de *clusters*, por consiguiente, se tiene una nueva modalidad de segmentación totalmente independiente a la antigüedad de la cartera, pues la misma se basa en otras variables que hacen posible focalizar la gestión de cobranza de mejor manera y permiten establecer estrategias diferenciadas.

Por último, al ser la distancia el discriminante en la formación de *clusters*, ésta puede ser el elemento que facilite el ordenamiento, pues prioriza la pertenencia a uno u otro *cluster*; por lo tanto, es factible usar este concepto como enfoque de prioridad y distribución para la gestión de cobranza.

Dentro de las limitaciones del estudio está la antigüedad de la información proveniente del buró, dado que la misma se obtuvo al otorgamiento del crédito. Por ello, para futuras investigaciones se recomienda actualizar la información crediticia de los clientes en base a su condición actual, además de la inclusión de la probabilidad de default en el modelo. De manera complementaria, se recomienda un estudio referente a nuevas estrategias de cobranza en función de nuevas agrupaciones y comportamientos; y a su vez, es indispensable profundizar el ordenamiento facilitado por la distancia o grado de atracción de los *clusters* al centroide.

## 5. Bibliografía

Chiu, C. Y., Chen, Y. F., Kuo, I. T., y Ku, H. C. (2009). *An intelligent market segmentation system using k-means and particle swarm optimization*. *Expert Systems with Applications*, 36(3 PART 1), 4558–4565. <https://doi.org/10.1016/j.eswa.2008.05.029>

Coronel Hoyos, M. A. (2019). *Diseño de un modelo para la gestión de cobranzas en la Cooperativa de Ahorro y Crédito Alianza del Valle*. Universidad Andina Simón Bolívar.

Dash M., Koot P.W. (2009) *Feature Selection for Clustering*. In: LIU L., ÖZSU M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-39940-9\\_613](https://doi.org/10.1007/978-0-387-39940-9_613)

Deloitte. (2012). *Tendencias de cobranza y recuperación de cartera en el sector financiero a partir de la crisis*. <https://www2.deloitte.com/content/dam/Deloitte/pa/Documents/financial-services/2015-01-Pa-FinancialServices-CobranzaCartera.pdf>

Equifax. (24 de agosto de 2020). Cerca de

3 millones de ecuatorianos podrían ingresar en vulnerabilidad económica a causa de la pandemia. *Equifax*.

[https://www.equifax.ec/acerca-de-equifax/comunicados-de-prensa/-/blogs/cerca-de-3-millones-de-ecuatorianos-podrian-ingresar-en-vulnerabilidadeconomica-a-causa-de-la-pandemia?\\_com\\_liferay\\_blogs\\_web\\_portlet\\_BlogsPortlet\\_redirect=https%3A%2F%2Fwww.equifax.ec%2Facerca-de-equifax%2Fcomunicados-de-prensa%3Fp\\_p\\_id%3Dcom\\_lifera\\_y\\_blogs\\_web\\_portlet\\_BlogsPortlet%26p\\_p\\_lifecycle%3D0%26p\\_p\\_state%3Dnormal%26p\\_p\\_mode%3Dview%26\\_com\\_liferay\\_blogs\\_web\\_portlet\\_BlogsPortlet\\_cur%3D1%26\\_com\\_liferay\\_blogs\\_web\\_portlet\\_BlogsPortlet\\_delta%3D75](https://www.equifax.ec/acerca-de-equifax/comunicados-de-prensa/-/blogs/cerca-de-3-millones-de-ecuatorianos-podrian-ingresar-en-vulnerabilidadeconomica-a-causa-de-la-pandemia?_com_liferay_blogs_web_portlet_BlogsPortlet_redirect=https%3A%2F%2Fwww.equifax.ec%2Facerca-de-equifax%2Fcomunicados-de-prensa%3Fp_p_id%3Dcom_lifera_y_blogs_web_portlet_BlogsPortlet%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26_com_liferay_blogs_web_portlet_BlogsPortlet_cur%3D1%26_com_liferay_blogs_web_portlet_BlogsPortlet_delta%3D75)

Equifax. (02 de septiembre de 2020). Las provincias de la Costa son las que más recurren a las compras a crédito en casas comerciales. *Equifax*.

<https://www.equifax.ec/acerca-de-equifax/comunicados-de-prensa/-/blogs/las-provincias-de-la-costa-son-las-que-mas-recurren-a-las-compras-a-credito-en-casas->

- comerciales?\_com\_liferay\_blogs\_web\_portlet\_BlogsPortlet\_redirect=https%3A%2F%2Fwww.equifax.ec%2Facerca-de-equifax%2Fcomunicados-de-prensa%3Fp\_p\_id%3Dcom\_liferay\_blogs\_web\_portlet\_BlogsPortlet%26p\_p\_lifecycle%3D0%26p\_p\_state%3Dnormal%26p\_p\_mode%3Dview%26\_com\_liferay\_blogs\_web\_portlet\_BlogsPortlet\_cur%3D1%26\_com\_liferay\_blogs\_web\_portlet\_BlogsPortlet\_delta%3D75
- Escobedo, M. T., y Salas, J. A. (2008). *P. Ch. Mahalanobis y las aplicaciones de su distancia estadística*. Culcyt, 27, 13-20.
- Jiang, Y., y Stylos, N. (2021) *Triggers of consumers' enhanced digital engagement and the role of digital technologies in transforming the retail ecosystem during COVID-19 pandemic*. Technological Forecasting and Social Change, 172(July), 121029. <https://doi.org/10.1016/j.techfore.2021.121029>
- Kennedy, K., Mac Namee, B., Delany, S. J., O'Sullivan, M., y Watson, N. (2013). *A window of opportunity: Assessing behavioural scoring*. Expert Systems with Applications, 40(4), 1372–1380. <https://doi.org/10.1016/j.eswa.2012.08.052>
- Microsoft. (11 de abril de 2006). *Component: K-Means Clustering*. <https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/k-means-clustering>
- Mingoti, S. A., y Lima, J. O. (2006). *Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms*. European Journal of Operational Research, 174(3), 1742–1759. <https://doi.org/10.1016/j.ejor.2005.03.039>
- Nanda, S. R., Mahanty, B., y Tiwari, M. K. (2010). *Clustering Indian stock market data for portfolio management*. Expert Systems with Applications, 37(12), 8793–8798. <https://doi.org/10.1016/j.eswa.2010.06.026>
- Riera Naranjo, B. (2018). *Desarrollo de un modelo de scoring de segmentación de cobranzas para tarjeta de crédito de la banca de personas de Produbanco*. Universidad Andina Simón Bolívar.

Rodriguez, A. (2018). *Modelo scoring para recuperar cartera de microcredito*. Fundación Universitaria Los Libertadores.

Shin, H. W., y Sohn, S. Y. (2004). *Segmentation of stock trading customers according to potential value*. *Expert Systems with Applications*, 27(1), 27–33. <https://doi.org/10.1016/j.eswa.2003.12.002>

Singh, A., Yadav, A., y Rana, A. (2013). *K-means with Three different Distance Metrics*. *International Journal of Computer Applications*. 67(10), 13-17. <https://doi.org/10.5120/11430-6785>

Wittlinger B., Carranza L., y Mori T. *“Mejores Prácticas en Estrategias de Cobranza”*, ACCION No. 26, ACCION Internacional Headquarters, Noviembre 2008.

Yang, W., Long, H., Ma, L., y Sun, H. (2020). *Research on clustering method based on weighted distance density and k-means*. *Procedia Computer Science*, 166, 507–511. <https://doi.org/10.1016/j.procs.2020.02.056>