



Modelo de puntuación crediticia para tarjeta de crédito en México: una aproximación logística

Credit Scoring Model for Credit Card in Mexico: A Logit Approach

Marco Antonio Reyes Morales^{§*}
Miriam Sosa^{*}

Información del artículo

Recibido:
23 septiembre 2021

Aceptado:
8 abril 2022

Clasificación JEL:
G21; G32; D81

Palabras clave:
Puntuación Crediticia;
Tarjeta de Crédito;
México; Modelo
Logístico

Resumen

El riesgo de crédito es una de las principales preocupaciones de los organismos de supervisión y regulación financiera, así como de las instituciones bancarias. Por lo que, se propone un modelo de puntuación crediticia basado en una regresión logística, para analizar la probabilidad de incumplimiento por segmentos de una cartera de clientes de tarjeta de crédito de una institución mexicana. Los resultados muestran que el modelo propuesto tiene un alto nivel de predictibilidad y de estabilidad, tanto fuera como dentro del periodo de modelado, la comprobación de monotonidad, también asegura que el modelo tenga un alto nivel de precisión. La originalidad subyace en que existen escasos estudios sobre modelos de puntuación crediticia para México, el resultado del modelo tiene alto nivel de precisión y arroja como resultado una tabla de puntuación de fácil interpretación para el personal bancario. Se concluye que el modelo es confiable y con alto nivel de ajuste.

* UNAM; [§] marco_antoniol125@hotmail.com



Article information	Abstract
Received: 23 September 2021	Credit risk is one of the main concerns of the financial institutions and supervision and regulation organisms. Thus, it is proposed a credit scoring model based on logit approach to analyze the default risk for a credit card portfolio in a Mexican financial institution. Findings show that the model proposed has a high level of prediction and stability, in and out of the sample. The monotonicity property evidences that the model has a high level of precision. The originality lies in the fact that, there is scarce literature on credit scoring models for Mexico. Results of the model are highly accurate in terms of predictability and the evidence is presented in a scoring table that is easy to interpret for all bank employees. We conclude that the model is reliable and highly accurate.
Accepted: 8 April 2022	
JEL Classification: G21; G32; D81	
Keywords: Credit Scoring; Credit Card; Mexico; Logit Model	

Introducción

Una de las principales actividades y negocios bancarios es la concesión de crédito. Sin embargo, esta actividad expone a las instituciones financieras a diversos tipos de riesgos, entre los que destacan el riesgo de crédito (Caruso, Gattone, Fortuna y Di Battista, 2021). El riesgo crediticio es aquella probabilidad de que el sujeto no cumpla con las obligaciones derivadas de una deuda (Bystrôm, 2019). Si una gran cantidad de sujetos incumplen con sus obligaciones, no solamente se incrementarán los costos financieros de la operación, sino que, también se generarán costos administrativos asociados a la recuperación de dichos créditos; afectando, los flujos de efectivo proyectados por dichas instituciones, pudiendo generar problemas de liquidez e insolvencia (Chuan y Lin, 2009).

En el contexto financiero global contemporáneo, caracterizado por una profunda interconexión entre instituciones y agentes localizados en diversos espacios geográficos, los problemas que se presenten en cierta entidad tenderán a esparcirse en el resto del sistema, incrementando el riesgo de un colapso generalizado, tal como ha ocurrido en episodios anteriores de crisis financieras: crisis mexicana de 1994; crisis asiática, 1997; crisis rusa, 1998; crisis brasileña, 1999; crisis dot com, 2001; crisis subprime, 2007 y crisis de la deuda soberana de 2012.

Ante la contingencia sanitaria generada por el COVID19 y el deterioro económico, causado por las medidas de prevención de contagio, cobra una especial importancia el análisis que las instituciones bancarias realizan sobre sus clientes, para disminuir la probabilidad de riesgo crediticio. No obstante, una sobreestimación del riesgo podría tener efectos importantes en la rentabilidad, la participación de mercado y la competitividad de los bancos. Por lo que, las instituciones financieras están sumamente interesadas en

detectar las necesidades del cliente, y ofrecerle productos de acuerdo con sus capacidades financieras (Papouškova y Hajek, 2019).

Socialmente y en el nivel macroeconómico, las decisiones de concesión de crédito son muy importantes, ya que expanden la capacidad de consumo de los individuos, y las de producción de las empresas; incrementando el nivel de actividad económica. Sin embargo, la expansión del crédito de manera irresponsable podría generar una sobrecarga financiera para los diversos agentes, derivando en fragilidad y vulnerabilidad locales, lo cual incrementaría el riesgo de una crisis bancaria.

Teóricamente, el riesgo crediticio se hace presente en una situación de información asimétrica, la cual hace referencia a que una de las contrapartes de un contrato (el banco en este caso) no posee suficiente información sobre la otra contraparte (el cliente), por lo que, se tiende a otorgar crédito a malos pagadores o malos proyectos y se niega crédito a aquellos sujetos/proyectos que tienen altas probabilidades de cumplir con sus compromisos (Vega, Santoyo, Muñoz y Altamirano, 2017). Para sobreponerse a dicha situación, las instituciones financieras han desarrollado y afinado técnicas de análisis de riesgo, para tratar de predecir el comportamiento de los individuos dado un conjunto de características.

Una de las técnicas más empleadas en la banca para mitigar el riesgo de crédito es el *credit scoring*, el cual consiste en un sistema de valoración crediticia que evalúa automáticamente el riesgo de cada solicitud de crédito (Puertas y Martí, 2012). Dicho riesgo se encuentra en función de un conjunto de variables asociadas al producto y al individuo. Los modelos automáticos de calificación crediticia optimizan el análisis de la información, disminuyendo los recursos asociados a dicha actividad, permitiendo análisis y decisión oportunos. Asimismo, tienen como propósito evitar la autorización de créditos a clientes que podrían ser incumplidos, lo cual ocasionaría grandes quebrantos para la institución concededora (Siddiqi, 2017).

Uno de los problemas que se presenta en los sistemas de calificación crediticia es el análisis del resultado para la toma de decisión, ya que requiere de transformar los parámetros de sensibilidad de cada variable β , en la probabilidad y la creación de una tabla de puntajes. Por lo que, en el presente trabajo, se propone la sistematización de dicha información, de forma que pueda ser entendida y empleada en las distintas áreas de la entidad.

Con base en lo previamente señalado, el objetivo de esta investigación es diseñar un modelo de puntuación del comportamiento crediticio, empleando una regresión logística binaria, la cual permite obtener mediante una ecuación, la probabilidad de realización de un evento (variable dependiente), dado un

conjunto de variables independientes. Se ha elegido dicho modelo por la economía en términos de los recursos necesarios para estimarlo, a diferencia de modelos más complejos que requieren mayor tiempo y capacidad de procesamiento. La principal contribución radica en que existen escasos estudios sobre modelos de puntuación crediticia para México. El resultado del modelo tiene alto nivel de precisión y arroja como resultado una tabla de puntuación de fácil interpretación para el personal bancario.

El modelo de puntuación crediticia transforma la probabilidad resultante de la regresión en un puntaje, el cual se interpreta como una calificación donde a mayor puntaje, menor el riesgo o menor la probabilidad asociada. Se utilizan datos del portafolio de tarjeta de crédito de una institución financiera mexicana de abril 2017 a marzo 2018, por cuestiones de confidencialidad a los datos se les aplica un factor solo para su presentación en la presente investigación.

La hipótesis que se desprende de la presente investigación es que los modelos de *credit scoring* son adecuados para la toma de decisiones de crédito debido a su alto nivel de precisión, a su alta predictibilidad y a su fácil entendimiento. La originalidad y contribución de la presente investigación subyacen en la propuesta de tabla de puntajes generada a partir del modelo y en la comprobación del ajuste de este, mediante una prueba fuera del periodo de construcción, lo cual asegura la robustez del modelo en cuanto a estabilidad en el tiempo, asegurando que el modelo califica cada cuenta del portafolio de tarjetas de crédito según su probabilidad de impago.

El trabajo se estructura en cinco partes, la segunda sección revisa la literatura relacionada, la tercera parte describe la metodología y datos empleados, la sección cuatro analiza los resultados y la última sección concluye la investigación.

1. Revisión de la literatura

El tema del riesgo crediticio y los indicadores asociados a su categorización, así como la toma de decisiones a partir de ellos, ha sido de sumo interés para las instituciones financieras, las autoridades económicas encargadas de la supervisión y regulación bancaria y para la sociedad en su conjunto.

La banca tiene un papel primordial en términos económicos, por las operaciones de financiamiento, pero también es la encargada de resguardar los ahorros de la sociedad, por lo que, su adecuado funcionamiento y la minimización de riesgos asociados a su operación han sido temas extensamente tratados por la literatura. Así que, existe una gran cantidad de estudios enfocados en predecir una situación de banca rota (Alaka, et al., 2018;

Davi y Radhika, 2018; Ansari, Ahmad, Bakar y Yaakub, 2020, y Zhang et al., 2021).

El análisis de riesgo de crédito ha sido desarrollado para organizaciones con diversos productos financieros. En términos de microcréditos, Ampountolas, Nyarko Nde y Constantinescu (2021) proponen un modelo de *credit-scoring* basado en aprendizaje automático, “machine learning”, sus resultados apuntan a que la metodología propuesta da resultados confiables, sobre todo, en ausencia de historiales crediticios o antecedentes de los clientes. De manera similar, Medina-Olivares, et al. (2021), desarrollan un modelo de *credit scoring* para microfinancieras, la originalidad de su propuesta radica en la inclusión de efectos espaciales, para así lograr mejorar el nivel de predicción del modelo.

En línea con el objeto de estudio del presente trabajo, Okesola et al. (2017), proponen un modelo basado en algoritmos bayesianos para evaluar el riesgo crediticio de tarjeta de crédito, empleando variables demográficas e indicadores materiales como variables de entrada. Los resultados muestran que el modelo y variables propuestas mejoran el resultado en rapidez, eficiencia y memoria efectiva.

Torvekar y Game (2019) aplican diversas técnicas de aprendizaje automático para identificar el comportamiento de los clientes, con el objetivo de predecir los clientes morosos de tarjeta de crédito. Asimismo, Ignatius et al. (2018), desarrollan una herramienta para la decisión sobre un modelo de puntuación de crédito basado en los principios de toma de decisiones multicriterio. Los resultados de la implementación son robustos y permiten la toma de decisión y adaptación del modelo ante diversas circunstancias.

En términos de economías emergentes, como es el caso de la economía mexicana y en la cual se basa la presente investigación, Trejo, Ríos y Martínez (2016) analizan la administración del riesgo crediticio en México para tarjetas de crédito. Implementan un modelo predictivo de incumplimiento en línea con la regulación local. La metodología empleada es un modelo logit y los resultados sugieren que provee un buen ajuste para la predicción de incumplimiento de clientes de tarjeta de crédito.

Trejo, et al. (2017), desarrollan una técnica de calificación crediticia, minimizando la pérdida esperada en el mercado de tarjeta de crédito a partir de una regresión logística. Los resultados apuntan a que, con el modelo propuesto, los costos de mantener créditos con niveles de riesgo mayores son más altos que para riesgos moderados.

Siguiendo la investigación realizada en México, el presente estudio propone una metodología de calificación crediticia basada en una regresión logística, empleando los datos de una cartera de clientes de una determinada institución bancaria. Dicho modelo permite predecir adecuadamente el incumplimiento de clientes de tarjetas de crédito, y transforma los resultados en indicadores que cualquier empleado de dicha institución, puede utilizar para la toma de decisiones.

2. Datos y metodología

2.1 Datos

Se utilizan datos del portafolio de tarjeta de crédito de una institución financiera mexicana de abril 2017 a marzo 2018; por cuestiones de confidencialidad, a los datos se les aplica un factor solo para su presentación en esta investigación. A dicha población, se le realizan ciertas exclusiones, las cuales se detallan en la tabla 1. Dichas exclusiones tienen el objetivo de realizar un modelado más apegado a la realidad y de desarrollar un modelo más limpio, sin sesgos en el comportamiento de las cuentas.

Tabla 1
Exclusiones

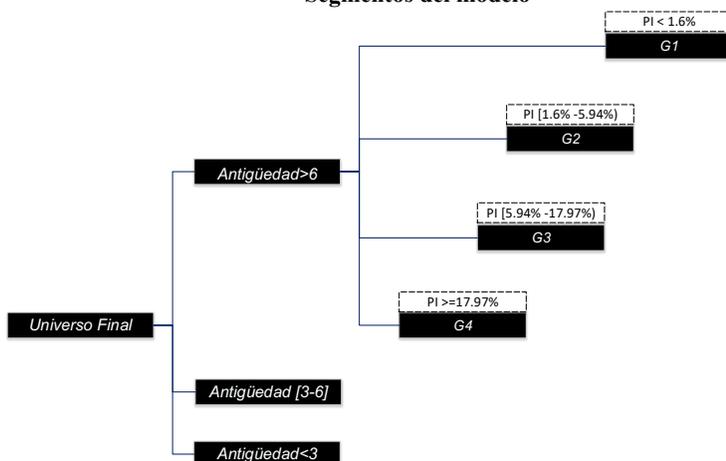
Decantación de cuentas	# Cuentas
Universo Inicial	27,795,663
Cuentas inactivas	-15,487,191
2 pagos vencidos o más ¹	-504,319
Universo Final	11,804,153

Fuente: elaboración propia.

Con el objetivo de incrementar la predictibilidad en el modelo, se segmenta el portafolio en grupos y subgrupos considerando características homogéneas entre las cuentas de cada uno de ellos. La segmentación se aprecia en el siguiente diagrama (figura 1), en donde se muestra que la primera variable que se segmenta es la antigüedad de la cuenta y la segunda, la probabilidad de incumplimiento (PI).

¹ Se excluyen las cuentas con impago, ya que su probabilidad de incumplir es muy cercana a uno.

Figura 1
Segmentos del modelo



Fuente: Elaboración propia.

Nota: Los rangos de la PI son determinados por la institución con base en la circular única de bancos.

Como se aprecia en el diagrama de la figura 1, la primera partición o grupo es por la antigüedad en meses que tienen las cuentas al momento de su calificación, es importante esta división ya que, la información de cada subconjunto es diferente debido a la disponibilidad y la precisión de tal información.

Por otro lado, la división por antigüedad en meses de la cuenta deja a la institución en incertidumbre del riesgo, ya que normalmente las cuentas prematuras conllevan un mayor riesgo que las cuentas con mayor maduración. Dado lo anterior, se crean tres grupos como se observa en el diagrama, el primero con cuentas con una antigüedad menor a tres meses, un segundo grupo con aquellas cuentas que tienen de tres a seis meses de antigüedad y, por último, el tercer grupo con cuentas que tienen más de seis meses de haber realizado su apertura.

La Comisión Nacional Bancaria y de Valores (CNBV) establece metodologías generales para la calificación de la cartera crediticia de las instituciones bancarias, con el fin de calcular las reservas que los bancos deben constituir. Dentro de estas metodologías, es obligación de la institución bancaria contar con una probabilidad de incumplimiento para cada cuenta de tarjeta de crédito. El cálculo de dicha Probabilidad de Incumplimiento (PI) es obligatoria e imprescindible para las instituciones financieras en México, por lo que se aprovecha dicha Probabilidad de Incumplimiento (PI) como variable para

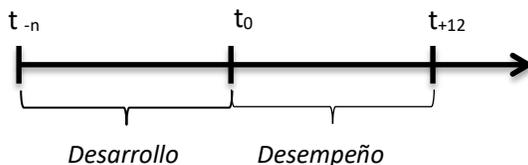
segmentar el presente modelo, con el objetivo de focalizar aún más la información a cada subgrupo de probabilidad.

Los subgrupos quedan como se muestra en el diagrama 1. El primer subgrupo, con una probabilidad menor al 1.6% de PI; un segundo grupo, con un rango de 1.6% a 5.94%; el tercer grupo, con una probabilidad de 5.94% a 17.97% y, un último grupo, con una probabilidad mayor a 17.97%.

La variable objetivo o variable dependiente se define como, toda aquella cuenta que haya incumplido 90 días o más en los 12 meses posteriores a su calificación, esta definición está alineada a las metodologías generales para la calificación de la cartera crediticia de las instituciones bancarias de la CNBV², así como a las recomendaciones de los acuerdos de Basilea II.

Para la construcción del modelo son utilizados dos periodos, el de Desarrollo y el de Desempeño, los cuales se muestran en el diagrama de la figura 2. Se parte de t_0 momento en el cual se califica con base en la información obtenida del periodo t_n a t_0 , con el objetivo de pronosticar los resultados del periodo t_0 a t_{+12} .

Figura 2
Periodos de información



Fuente: Elaboración propia.

El periodo de desarrollo (histórico), se refiere a la información sobre el comportamiento y las características de las cuentas en n -meses previos a la fecha de referencia (t_0). Dicha información se analiza con el fin de generar variables históricas que pudieran explicar mejor la variable objetivo.

Periodo de desempeño (observación), es el intervalo de doce meses posteriores a la fecha de referencia (t_0). El propósito es comprobar el cumplimiento de la variable objetivo de cada cuenta, en donde toma el valor de 1 cuando la cuenta

² En las disposiciones de carácter general, aplicables a las instituciones de crédito de la CNBV, se especifica que todos los créditos en etapa 3 son aquellos que presentan un atraso mayor o igual que 90 días.

haya incumplido 90 días o más en dicho periodo y en caso contrario, se le asigna el valor de 0.

Se construyen 131 variables con base en el periodo de desarrollo, en la tabla 2 se muestra la clasificación de dichas variables y el número de estas, consideradas en cada categoría. Dichas categorías se explican a continuación.

Tabla 2
VARIABLES CONSTRUIDAS

Categoría	Número de variables
Información de comportamiento interno	51 variables
Información de apertura	34 variables
Información de comportamiento externo	42 variables
Sociodemográfica	4 variables

Fuente: Elaboración propia.

Información de comportamiento interno. -Se refiere al comportamiento de las propias cuentas de tarjeta de crédito en cada cierre de mes, es decir, son variables propias del manejo del crédito, como pueden ser: su porcentaje de uso, el porcentaje de pago, moras históricas, perfil de la tarjeta de crédito (clásica, oro, premium), pagos requeridos, etc.

Información de apertura. -Este tipo de variables, se construye tomando los datos que el cliente otorga al momento de solicitar la apertura de la cuenta de tarjeta de crédito, dicha información puede ser: monto de sus ingresos, créditos vigentes, porcentaje de uso de créditos vigentes, etc.

Información de comportamiento externo. -Esta información se refiere al comportamiento que el cliente muestra, sobre otros créditos en el mercado, como, por ejemplo: porcentaje de uso de todas sus tarjetas de crédito, suma del total de sus pagos requeridos de los créditos hipotecarios, morosidades en sus créditos de nómina, entre otros.

Información sociodemográfica. -Es la información actualizada que el cliente otorga a la institución referente a su edad, sexo, lugar de referencia, etc.

3. Metodología

La técnica con la que se realiza el modelo es una regresión logística binaria, la cual permite obtener mediante la ecuación (1), la probabilidad de ocurrencia de un evento (variable dependiente) dado un conjunto de variables independientes. El modelo de *credit scoring* transforma la probabilidad resultante de la regresión en un puntaje, el cual se interpreta como una calificación donde a mayor puntaje, menor el riesgo o menor la probabilidad asociada.

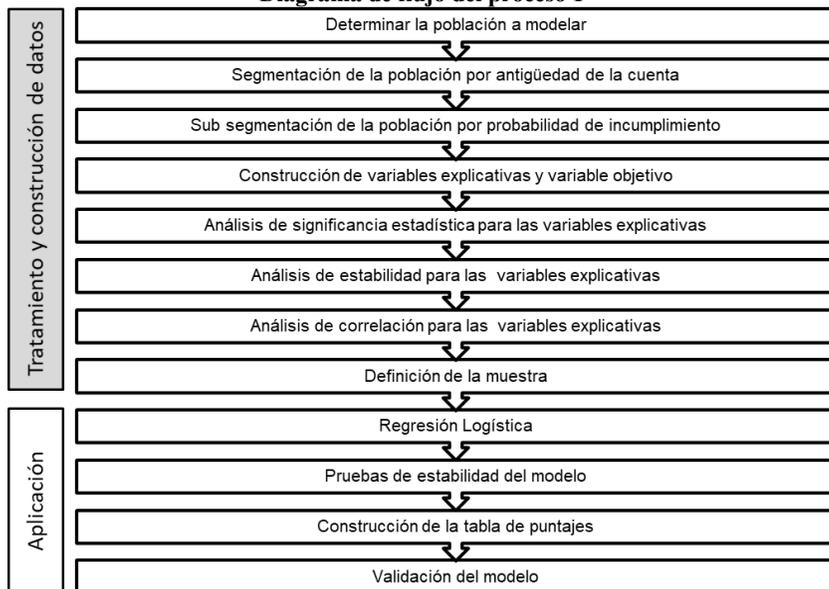
“El modelo logístico establece la siguiente relación entre la probabilidad de que ocurra el suceso, dado que el individuo presenta los valores $(X = x_1, X = x_2, \dots, X = x_k)$:

$$P[Y = 1 / x_1, x_2, \dots, x_k] = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad (1)$$

El objetivo es hallar los coeficientes $(\beta_0, \beta_1, \dots, \beta_k)$, que mejor se ajusten a la expresión funcional”. (De la Fuente, 2011, p.7)

En el diagrama de flujo del proceso de la figura 3, se explica detalladamente los pasos de la metodología aplicada.

Figura 3
Diagrama de flujo del proceso 1



Fuente: Elaboración propia.

4. Análisis de los resultados

Con el fin de conocer los estadísticos principales, tales como: frecuencia total, frecuencia de nulos, media, moda, mediana, desviación estándar, percentiles, máximo y mínimo, las variables son sometidas a un análisis univariado (véase apéndice 1). Lo anterior permite, principalmente, identificar la tasa de ausentes y con esto definir su posterior tratamiento en caso de ser necesario.

Para encontrar las variables que mejor explican la variable objetivo (que la cuenta en un periodo de 12 meses haya caído en 90 o más días de impago), se realiza un análisis de significancia estadística. Dentro de dicho análisis se realizan las siguientes pruebas: Índice de GINI, Information Value y Prueba de Kolmogorov–Smirnov.

Dado el análisis de significancia estadística se determina que, de las 131 variables, solo 22 explican estadísticamente a la variable objetivo, tienen un *p-value* menor a 0.0001, (véase apéndice 2 para la descripción de cada una de estas).

En las siguientes tablas (3 y 4) se muestran las variables con mayor significancia de cada segmento y sus atributos, así como el resultado del análisis estadístico de cada variable.

Tabla 3
Variables explicativas Segmento antigüedad <3 meses

Segmento	Variable	Rango	Tasa malos	GINI	KS	IV
Antigüedad < 3 meses	uso_interno	< 0.10	5.87%	39.39%	35.94%	0.76
		[0.10 - 0.45]	14.61%			
		[0.45 - 0.95]	21.35%			
	score_apertura	>= 0.95	46.08%	39.25%	28.99%	0.58
		< 368	21.95%			
		[368 - 385]	11.26%			
		[385 - 398]	5.69%			
	prom_pagoexige_6m	>= 398	2.14%	26.52%	24.94%	0.47
		< 0.057	7.48%			
		[0.057 - 0.2268]	14.07%			
		[0.2268 - 0.5669]	24.43%			
	ant_max_tdc_creacion	>= 0.5669	42.30%	24.73%	22.18%	0.23
		sin información	14.94%			
		< 25	17.02%			
		[25 - 35]	11.02%			
		[35 - 59]	8.34%			
	util_tdc_creacion	>= 59	5.97%	24.44%	21.11%	0.24
		sin información	13.31%			
		<= 0.34	4.65%			
		[0.34 - 0.69]	7.41%			
		[0.69 - 0.95]	12.39%			
		>= 0.95	20.02%			

Fuente: Elaboración propia.

Como se observa, en el primer segmento del modelo referente a las cuentas de recién apertura, se tienen variables explicativas basadas en información interna y, sobre todo, en variables de apertura. Lo anterior tiene sentido, ya que han transcurrido a lo más dos meses desde que la cuenta de tarjeta de crédito fue abierta.

Para el segmento de antigüedad de entre tres y seis meses de antigüedad, las variables con mayor predicción se muestran en la tabla 4.

Tabla 4
Variables explicativas: Segmento antigüedad 3-6 meses

Segmento	Variable	Rango	Tasa malos	GINI	KS	IV
Antigüedad 3-6 meses	prom_uso_3m	< 0.15	2.81%	57.98%	20.02%	0.51
		[0.15 - 0.45]	12.83%			
		[0.45 - 0.8]	22.51%			
		>= 0.8	47.24%			
	prom_pago_6m	sin información	45.86%	53.92%	20.02%	0.38
		[0 - 0.1]	29.11%			
		[0.1 - 0.4]	19.45%			
		[0.4 - 0.75]	7.86%			
	veces_premora_6m	>= 0.75	3.82%	40.18%	13.31%	0.67
		< 1	5.95%			
		[1 - 2]	25.12%			
		[2 - 3]	33.64%			
	score_apertura	>= 3	41.67%	35.57%	17.02%	0.52
		<= 367	20.74%			
		[367 - 384]	11.70%			
		[384 - 401]	5.95%			
	util_tdc_creacion	>= 401	2.11%	21.88%	42.30%	0.21
		sin información	4.89%			
		<= 0.34	8.28%			
		[0.34 - 0.7]	12.74%			
	ant_max_tdc_creacion	[0.7 - 0.95]	13.48%	21.69%	42.30%	0.18
		>= 0.95	19.71%			
		sin información	16.77%			
		<= 24	14.17%			
	[24 - 36]	11.76%				
	[36 - 54]	9.25%				
		>= 54	6.58%			

Fuente: Elaboración propia.

Como se aprecia en la tabla anterior, las variables que pueden explicar con mayor exactitud a la variable objetivo son variables con información de creación, así como de información histórica. Dado que en este segmento ya se cuenta con más información interna, las variables explicativas son promedios de a lo más seis meses de información de desarrollo, esto permite discriminar mejor las cuentas que presentan una mayor probabilidad de evento.

Como se explica, para los segmentos cuya antigüedad es mayor a seis meses, se realizan cuatro subsegmentos por la probabilidad incumplimiento regulatoria de la CNBV. Las variables explicativas con mayor predictibilidad se muestran a continuación.

Tabla 5
Variables explicativas: Segmento antigüedad >6 meses Grupo 1

Segmento	Variable	Rango	Tasa malos	GINI	KS	IV
Antigüedad > 6 Probabilidad de incumplimiento G1	tipo_marca_tarjeta	CLASICA	1.43%	15.81%	12.79%	0.09
		ORO	2.02%			
		COMPARTIDA	0.99%			
		PREMIUM	0.77%			
		PLATINO	0.81%			
	cuenta_cheques_interna	NO	2.11%	14.81%	14.81%	0.08
		SI	1.19%			
	prom_pago_3m	< 0.35	1.98%	19.85%	19.51%	0.16
		[0.35 - 0.75]	1.18%			
		>= 0.75	0.79%			
	Veces_mora_externa_12M	< 3	2.67%	16.51%	15.50%	0.12
		[3 - 9]	1.83%			
>= 9		1.20%				

Fuente: Elaboración propia.

Tabla 6
Variables explicativas: Segmento antigüedad >6 meses Grupo 2

Segmento	Variable	Rango	Tasa malos	GINI	KS	IV
Antigüedad > 6 Probabilidad de incumplimiento G2	perfilamiento_pago	Baja morosidad	6.60%	16.86%	13.80%	0.16
		Alta morosidad	17.20%			
		Alta revolvencia	7.33%			
		Media revolvencia	6.33%			
		Baja revolvencia	6.22%			
		Totalero comisión	4.25%			
	prom_pago_6m	Totalero	1.60%	20.41%	18.72%	0.16
		<= 0.15	8.10%			
		[0.15 - 0.75]	7.00%			
	veces_premora_6m	>= 0.75	3.30%	15.64%	14.53%	0.10
		<= 1	4.55%			
		[1 - 2]	6.96%			
		[2 - 4]	8.56%			
	Veces_mora_externa_12M	>= 4	10.84%	12.71%	11.52%	0.06
		<= 2	7.66%			
		[2 - 6]	6.22%			
		[6 - 9]	5.31%			
	incr_pagoexig_6m	>= 9	4.57%	11.08%	9.15%	0.06
		< 2	4.81%			
		[2 - 3]	5.47%			
[3 - 5]		7.69%				
prom_uso_9m	>= 5	13.10%	9.64%	9.64%	0.04	
	< 0.63	4.87%				
dec_pago_9m	>= 0.63	7.28%	8.12%	7.52%	0.02	
	< 2	4.80%				
	[2 - 4]	6.23%				
	[4 - 6]	6.78%				
		>= 6	8.92%			

Fuente: Elaboración propia.

Para el primer subsegmento de probabilidad de incumplimiento, menor a 1.6%, se busca principalmente, separar a las cuentas con un comportamiento excelente de las cuentas buenas. Como se observa en la tabla 5, la mayoría de las variables son internas y se enfocan en la naturaleza del cliente y su manejo con la cuenta en el corto plazo.

En lo que respecta al segundo grupo de probabilidad de incumplimiento, de 1.6% a 5.94%, el objetivo es encontrar las variables explicativas que separan las cuentas buenas de las cuentas con mayor probabilidad a cumplir el evento. Estas se muestran en la tabla 6.

En el tercer subsegmento de PI, de 5.94% a 17.97%, ya se presenta mayor volumen de cuentas malas, por lo que el propósito de este subsegmento es

hallar las variables explicativas que separen las dichas cuentas de las cuentas que presentan mejor comportamiento y menor probabilidad de evento. Los resultados se muestran en la tabla 7.

Tabla 7
VARIABLES EXPLICATIVAS: Segmento antigüedad >6 meses Grupo 3

Segmento	Variable	Rango	Tasa malos	GINI	KS	IV
Antigüedad > 6 Probabilidad de incumplimiento G3	pct_ctasmora_externas	sin información	13.97%	12.71%	11.52%	0.04
		< 0.25	16.12%			
		[0.25 - 0.75)	20.72%			
		>= 0.75	26.82%			
	Ant_cuenta	< 24	24.93%	15.64%	14.53%	0.14
		>= 24	13.05%			
	veces_premora_6m	< 2	13.30%	12.71%	11.52%	0.05
		[2 - 4)	18.64%			
		>= 4	20.87%			
	veces_mora_externa_12M	< 3	18.74%	15.64%	14.53%	0.03
		[3 - 9)	14.06%			
		>= 9	14.12%			
	incr_pagoexig_6m	< 2	13.85%	11.08%	9.15%	0.07
		[2 - 4)	18.61%			
		>= 4	25.48%			
dec_pago_6m	< 2	14.90%	8.12%	7.52%	0.02	
	[2 - 3)	18.80%				
	[3 - 4)	19.50%				
	>= 4	20.70%				

Fuente: Elaboración propia.

Tabla 8
VARIABLES EXPLICATIVAS: Segmento antigüedad >6 meses Grupo 4

Segmento	Variable	Rango	Tasa malos	GINI	KS	IV
Antigüedad > 6 Probabilidad de incumplimiento G4	última_mora	Corriente	35.52%	25.39%	25.39%	0.41
		1 Pago vencido	74.12%			
	incr_uso_6m	< 2	32.80%	22.99%	18.54%	0.19
		[2 - 3)	44.40%			
		[3 - 5)	55.80%			
		>= 5	69.70%			
	ant_cuenta	< 24	56.81%	20.32%	20.32%	0.18
		>= 24	35.33%			
	veces_mora1_12m	< 1	26.27%	14.17%	10.56%	0.10
		[1 - 3)	43.82%			
		>= 3	48.96%			
	incr_pagoexig_6m	< 2	39.10%	12.69%	12.34%	0.09
		[2 - 3)	53.80%			
>= 3		62.50%				

Fuente: Elaboración propia.

Las variables que explican de mejor forma el grupo 3, como se muestra en la tabla anterior, es una mezcla de variables basadas en información interna y variables construidas a partir de información externa. Además, toman mayor relevancia las variables que contienen información histórica de demoras.

Para el último grupo de PI, mayor a 17.97%, se busca identificar mediante las variables, a las cuentas que podrían no presentar el evento, es decir a las cuentas buenas que el modelo de la PI está considerando con una probabilidad de incumplimiento alta. Las variables resultantes se muestran en la tabla 8.

Como se observa, la información de las variables explicativas para este subsegmento es información interna y externa de la cuenta, así como de últimas

demoras, ya que se espera que una cuenta morosa, tenga mayor probabilidad de presentar el evento.

Dado el análisis anterior, se puede determinar que todas las variables que se hallaron cumplen con los requisitos estadísticos para formar parte del modelo, para así poder predecir con mayor exactitud el evento o no evento en el periodo de desempeño.

Las variables que resultaron estadísticamente significativas son sometidas a un análisis de estabilidad, con el fin de darle rigor a la regresión.

Mediante el índice de estabilidad de la población (PSI), que indica la magnitud de cambio de las variables, es decir, la medida de cambio entre la población actual y la esperada. La manera de calcularlo se muestra en la siguiente fórmula.

$$\sum(\% \text{ Actual} - \% \text{ Esperado}) * \ln\left(\frac{\% \text{ Actual}}{\% \text{ Esperado}}\right) \quad (2)$$

En general, un índice menor a 0.1 es un cambio no significativo, un índice que toma un valor entre 0.10 y 0.25 indica que existe un pequeño cambio y un índice mayor a 0.25 indica que existe un cambio significativo para el cual se debe realizar un análisis más profundo, para la aceptación o rechazo de la variable. El resultado deseable es que el índice sea menor a 0.1, ya que indica que no existen cambios abruptos en la variable de un mes a otro.

Para el modelo propuesto, todas las variables se encuentran dentro del umbral menor a 0.1, por lo que estas no tienen un cambio significativo en el periodo de modelado y por lo tanto son aptas para formar parte del modelo a construir.

Con el objetivo de reducir la dimensionalidad³ se analizan los coeficientes, de dependencia o de correlación, de Pearson y Spearman, excluyendo aquellas variables en las que existiera un alto coeficiente. Ninguna de las 22 variables se descarta por dimensionalidad, ya que todas tienen como índice de correlación menor que 0.50.

Como se menciona, los datos del universo a modelar son más de 11 millones de cuentas, en un periodo que abarca de abril 2017 a marzo 2018. Dado el gran volumen de este, y con el objetivo de optimizar el manejo de la información, para el desarrollo del modelo, se realiza un muestreo aleatorio simple sin reposición.

³ Técnica de preprocesado para modelos probabilísticos cuyo objetivo es la reducción del número de variables de una colección de datos.

Dicho muestreo, para el caso de estudio, se realiza para cada segmento o subsegmento y para cada mes, con el propósito de modelar los impactos de temporalidad que pudieran existir en el universo.

Con el propósito de validar el modelo, al momento de desarrollarlo, se realizan dos submuestras, una de entrenamiento (*training data*) y una de validación (*test*), con el 70% y el 30% respectivamente del total de la muestra, coincidiendo con otros estudios como el realizado por Siddiqui (2006). La submuestra de entrenamiento se utiliza para el desarrollo propio modelado, mientras que la de validación, se utiliza para evaluar el modelado, comparando los resultados de las dos submuestras. Dichos resultados se explican en los siguientes capítulos.

Para el modelado *credit scoring*, se utiliza la técnica paramétrica Modelo logit o regresión logística, el objetivo de esta es otorgar una probabilidad de que ocurra la variable objetivo a cada cuenta del portafolio, asignando una ponderación (β) a cada atributo de cada variable independiente (X_1), y agregando un valor constante o intercepto a la ecuación (β_0), como se muestra en la fórmula 3 de la regresión logística.

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta X_1)}} \quad (3)$$

A cada uno de los segmentos y subsegmentos del modelo se le aplica una regresión logística, con el fin de que las probabilidades obtenidas sean equiparables entre estos, y así identificar las cuentas con mayor posibilidad de representar un incumplimiento en el futuro independientemente del segmento o subsegmento al que pertenezca.

En las siguientes tablas (9, 10 y 11), se muestran los pesos (β) de los atributos de cada variable, así como el intercepto o el valor constante (β_0) de cada regresión de cada uno de los segmentos propuestos⁴.

Como se aprecia en la tabla 9, la regresión logística del primer segmento arroja valores adecuados para las β , ya que entre mayor es la tasa de malos⁵ mayor es el valor de β , esto da la idea que, al momento de aplicar la fórmula de la regresión se obtendrán probabilidades más altas, las cuales serán coherentes con el comportamiento de la cuenta.

⁴ En los modelos de *credit scoring* toman relevancia los pesos (β) de los atributos de cada variable, así como las probabilidades obtenidas, a diferencia de otros modelos las pruebas de bondad de ajuste generalmente no son estimados, véase: López, P. P. (2018); Capelo Vinza, J. A. (2012); Gutierrez Girault, M. A. (2007); Cavallo, A., Coghlan, M., & Jackson, J. (2012).

⁵ Se refiere a la proporción de cuentas que cumplen con la variable objetivo (90+ días de impago en 12 meses) del total de cuentas que se encuentran en ese rango o atributo de la variable.

Por otro lado, como se observa en la tabla 9, el último atributo no arroja un valor para β , dado que en el cálculo de la regresión se toma el valor de la suma de las demás β por -1 de los demás atributos.

Para un mejor entendimiento de lo antes expuesto, se toma una cuenta de tarjeta de crédito, la cual tiene dos meses de haberse originado, por lo que la evaluación se realiza con la regresión logística de este segmento y en consecuencia se aplican los valores dados por la misma.

Por otro lado, partiendo de que la misma cuenta presenta los siguientes atributos:

- Uso_interno = 0.15
- Score_apertura = 400
- prom_pagoexige_6m = 0.25
- ant_max_tdc_creacion = 40
- util_tdc_creacion = 0.70

Tabla 9
Resultados regresión logística: Segmento antigüedad <3 meses

Segmento	Variable	Rango	Tasa malos	β
Antigüedad < 3 meses	uso_interno	< 0.10	5.87%	-1.0371
		[0.10 - 0.45]	14.61%	-0.0186
		[0.45 - 0.95]	21.35%	0.1486
	score_apertura	>= 0.95	46.08%	
		< 368	21.95%	1.0014
		[368 - 385]	11.26%	0.3881
		[385 - 398]	5.69%	-0.2470
	prom_pagoexige_6m	>= 398	2.14%	
		< 0.057	7.48%	-0.0944
		[0.057 - 0.2268]	14.07%	-0.2798
		[0.2268 - 0.5669]	24.43%	0.0127
	ant_max_tdc_creacion	>= 0.5669	42.30%	
		sin información	14.94%	0.0807
		< 25	17.02%	0.3212
		[25 - 35]	11.02%	0.0293
	util_tdc_creacion	[35 - 59]	8.34%	-0.1511
		>= 59	5.97%	
		sin información	13.31%	0.0479
		<= 0.34	4.65%	-0.3954
		[0.34 - 0.69]	7.41%	-0.1648
	[0.69 - 0.95]	12.39%	0.0798	
	>= 0.95	20.02%		
Intercepto				-1.7799

Fuente: Elaboración propia.

Los cálculos que realiza la regresión para encontrar la probabilidad son los siguientes.

Primeramente, se toma el valor de las β que corresponden a cada atributo de cada variable que presenta la cuenta, los cuales son los siguientes.

- Uso_interno = 0.15 → -0.0186
- prom_pagoexige_6m = 0.25 → 0.0126
- ant_max_tdc_creacion = 40 → -0.1510
- 34rea_tdc_creacion = 0.70 → 0.0797

Como se menciona, en el caso de que el atributo de la variable no tenga valor en la β , este se obtiene sumando los valores de las demás β y multiplicándolo por -1, es decir, para el caso de la cuenta evaluada, el cálculo se muestra como sigue:

$$(1.0014 + 0.3881 - 0.2470) * -1 = - 1. 1425 \tag{4}$$

Como se muestra en la tabla 10, los valores finales para la cuenta evaluada se encuentran sombreados en la columna con el nombre de Cálculo. Posteriormente, se suman todos los valores de las β y el valor del intercepto β_0 , como se aprecia en la tabla 10, en la fila con el nombre de Suma de Cálculos.

Tabla 10
Resultados regresión logística: Segmento antigüedad <3 meses

Segmento	Variable	Rango	β	Cálculo
Antigüedad < 3 meses	uso_interno	< 0.10	-1.0371	
		[0.10 - 0.45)	-0.0186	- 0.0186
		[0.45 - 0.95)	0.1486	
		>= 0.95		
	score_apertura	< 368	1.0014	
		[368 - 385)	0.3881	
		[385 - 398)	-0.2470	
		>= 398		- 1.1425
	prom_pagoexige_6m	< 0.057	-0.0944	
		[0.057 - 0.2268)	-0.2798	
		[0.2268 - 0.5669)	0.0127	0.0127
		>= 0.5669		
	ant_max_tdc_creacion	sin información	0.0807	
		< 25	0.3212	
		[25 - 35)	0.0293	
		[35 - 59)	-0.1511	- 0.1511
	util_tdc_creacion	>= 59		
		sin información	0.0479	
		<= 0.34	-0.3954	
		[0.34 - 0.69)	-0.1648	
	[0.69 - 0.95)	0.0798	0.0798	
	>= 0.95			
Intercepto				- 1.7799
Suma de cálculos				- 2.9997

Fuente: Elaboración propia.

Sustituyendo la suma de los valores de las β_1 calculadas, así como el intercepto en la fórmula 3, se tiene lo siguiente.

$$\frac{1}{1+e^{-(-2.9997)}} = 0.0474 \quad (5)$$

Como se observa, el resultado de la evaluación de la cuenta es 0.0474, el cual representa que dicha cuenta tiene una probabilidad de 4.74% de caer en 90+ días de impago en 12 meses posteriores a su evaluación. Los valores arrojados de la regresión para las cuentas con una antigüedad de 3 a 6 meses se muestran en la siguiente tabla.

Tabla 11
Resultados regresión logística: Segmento antigüedad 3-6 meses

Segmento	Variable	Rango	Tasa malos	β
Antigüedad 3-6 meses	prom_uso_3m	< 0.15	2.81%	-1.7709
		[0.15 - 0.45]	12.83%	0.1773
		[0.45 - 0.8]	22.51%	0.5409
		>= 0.8	47.24%	
	prom_pago_6m	sin información	45.86%	0.3352
		[0 - 0.1]	29.11%	0.5509
		[0.1 - 0.4]	19.45%	0.0462
		[0.4 - 0.75]	7.86%	-0.1776
	veces_premora_6m	>= 0.75	3.82%	
		< 1	5.95%	-0.5422
		[1 - 2]	25.12%	0.0831
		[2 - 3]	33.64%	0.1971
	score_apertura	>= 3	41.67%	
		<= 367	20.74%	0.8349
		[367 - 384]	11.70%	0.3470
		[384 - 401]	5.95%	-0.2468
	util_tdc_creacion	>= 401	2.11%	
		sin información	4.89%	0.0748
		<= 0.34	8.28%	-0.0732
		[0.34 - 0.7]	12.74%	-0.0145
	ant_max_tdc_creacion	[0.7 - 0.95]	13.48%	-0.0177
		>= 0.95	19.71%	
		sin información	16.77%	0.0729
		<= 24	14.17%	0.1779
Intercepto	[24 - 36]	11.76%	0.0923	
	[36 - 54]	9.25%	-0.0592	
	>= 54	6.58%		
	Intercepto			-1.6848

Fuente: Elaboración propia.

Nota: La significancia estadística de las betas, tiene un p-value menor a 0.0001.

En las siguientes tablas (12, 13, 14 y 15) se muestran los resultados que arroja cada regresión de cada subsegmento del segmento de antigüedad mayor a seis meses.

Tabla 12
Resultados regresión logística: Segmento antigüedad >6 meses Grupo 1

Segmento	Variable	Rango	Tasa malos	β	
Antigüedad > 6 Probabilidad de incumplimiento G1	tipo_marca_tarjeta	CLASICA	1.43%	0.2479	
		ORO	2.02%	0.5700	
		COMPARTIDA	0.99%	-0.5708	
		PREMIUM	0.77%	-0.1804	
		PLATINO	0.81%		
	cuenta_cheques_interna	NO		2.11%	0.2308
		SI		1.19%	
	prom_pago_3m	< 0.35		1.98%	0.4787
		[0.35 - 0.75]		1.18%	-0.1357
		>= 0.75		0.79%	
	Veces_mora_externa_12M	< 3		2.67%	0.2699
		[3 - 9]		1.83%	0.0665
		>= 9		1.20%	
	Intercepto				-4.4937

Fuente: Elaboración propia.

En cuanto a los resultados de la regresión para el grupo 2 del segmento antigüedad mayor a 6, se muestran en la tabla 13.

Tabla 13
Resultados regresión logística: Segmento antigüedad >6 meses Grupo 2

Segmento	Variable	Rango	Tasa malos	β	
Antigüedad > 6 Probabilidad de incumplimiento G2	perfilamiento_pago	Baja morosidad	6.60%	0.4072	
		Alta morosidad	17.20%	-0.5978	
		Alta revolvencia	7.33%	-0.2076	
		Media revolvencia	6.33%	-0.0268	
		Baja revolvencia	6.22%	-0.0865	
		Totalero comisión	4.25%	-0.8107	
		Totalero	1.60%		
	prom_pago_6m	<= 0.15		8.10%	0.5372
		[0.15 - 0.75]		7.00%	-0.0344
		>= 0.75		3.30%	
	veces_premora_6m	<= 1		4.55%	-0.4350
		[1 - 2]		6.96%	-0.0452
		[2 - 4]		8.56%	0.1305
		>= 4		10.84%	
	Veces_mora_externa_12M	<= 2		7.66%	0.2560
		[2 - 6]		6.22%	0.0177
		[6 - 9]		5.31%	-0.1205
		>= 9		4.57%	
	incr_pagoexig_6m	< 2		4.81%	-0.4968
		[2 - 3]		5.47%	-0.2058
		[3 - 5]		7.69%	0.1053
		>= 5		13.10%	
	prom_uso_9m	< 0.63		4.87%	-0.1652
		>= 0.63		7.28%	
	dec_pago_9m	< 2		4.80%	-0.0068
		[2 - 4]		6.23%	-0.0486
		[4 - 6]		6.78%	-0.1159
		>= 6		8.92%	
	Intercepto				-2.0206

Fuente: Elaboración propia.

En la siguiente tabla, se muestran los resultados de la regresión para cada atributo, así como el intercepto o la constante de esta, para el subsegmento de probabilidad de incumplimiento de 5.94% a 17.97% del segmento antigüedad mayor a 6.

Tabla 14
Resultados regresión logística: Segmento antigüedad >6 meses Grupo 3

Segmento	Variable	Rango	Tasa malos	β	
Antigüedad > 6 Probabilidad de incumplimiento G3	pct_ctasmora_externas	sin información	13.97%	-0.3332	
		< 0.25	16.12%	-0.1893	
		[0.25 - 0.75]	20.72%	0.0076	
		>= 0.75	26.82%		
	Ant_cuenta	< 24	24.93%	0.3671	
		>= 24	13.05%		
	veces_premora_6m	< 2	13.30%	-0.2899	
		[2 - 4]	18.64%	0.0499	
		>= 4	20.87%		
	veces_mora_externa_12M	< 3	18.74%	0.1360	
		[3 - 9]	14.06%	-0.1335	
		>= 9	14.12%		
	incr_pagoexig_6m	< 2	13.85%	-0.2721	
		[2 - 4]	18.61%	-0.0374	
		>= 4	25.48%		
	dec_pago_6m	< 2	14.90%	-0.0450	
[2 - 3]		18.80%	0.0089		
[3 - 4]		19.50%	-0.0085		
>= 4		20.70%			
Intercepto				-1.1995	

Fuente: Elaboración propia.

En lo que respecta al último subsegmento del segmento antigüedad mayor a 6 meses, los valores arrojados de la regresión se muestran en la tabla 15.

Tabla 15
Resultados regresión logística: Segmento antigüedad >6 meses Grupo 4

Segmento	Variable	Rango	Tasa malos	β
Antigüedad > 6 Probabilidad de incumplimiento G4	última_mora	Corriente	35.52%	-0.7612
		1 Pago vencido	74.12%	
	incr_uso_6m	< 2	32.80%	-0.4074
		[2 - 3]	44.40%	-0.1336
		[3 - 5]	55.80%	0.0655
		>= 5	69.70%	
	ant_cuenta	< 24	56.81%	0.4049
		>= 24	35.33%	
	veces_mora1_12m	< 1	26.27%	-0.2587
		[1 - 3]	43.82%	0.0586
		>= 3	48.96%	
	incr_pagoexig_6m	< 2	39.10%	-0.2406
		[2 - 3]	53.80%	0.0509
>= 3		62.50%		
Intercepto				0.5655

Fuente: Elaboración propia.

Como se observa, para todos los segmentos, los valores de β de cada atributo son directamente proporcionales a la tasa de malos que presenta cada uno de estos. Al igual que en los segmentos y subsegmentos anteriores, el resultado de la probabilidad depende de los atributos que tenga cada cuenta, ya que los valores de β de cada atributo son sustituidos, junto con el intercepto, en la fórmula de regresión logística para la obtención de esta.

Se concluye que las regresiones obtenidas para los distintos segmentos y subsegmentos son correctas, en cuanto a la asignación de valores a β de cada

uno de estos. Dichos valores de β son de relevancia para la construcción de la tabla de puntajes y del modelo de *credit scoring* en su totalidad, que es el objetivo final de la presente investigación.

Una vez obtenidos los valores de todas las regresiones, de todos los segmentos y subsegmentos, se procede a calificar con dichas regresiones a todas las cuentas del portafolio de tarjeta de crédito. Lo anterior con el objetivo de validar, primero, que las muestras que se consideraron para el modelado hayan sido representativas de la población y, por otro lado, que las probabilidades asignadas a cada una de las cuentas sean coherentes con el desempeño real de las mismas.

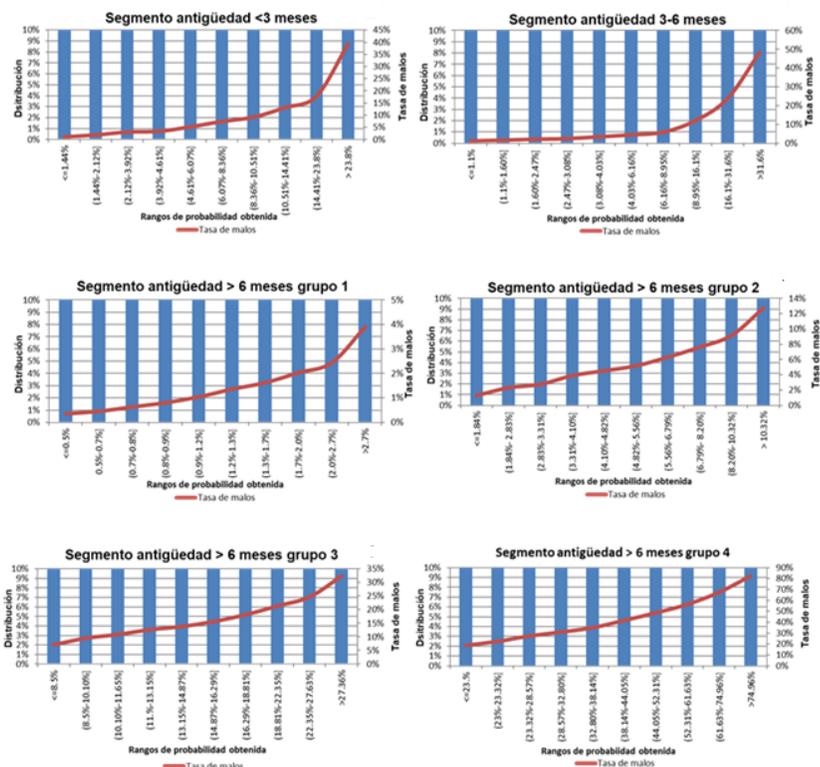
En la gráfica 1, se muestran las probabilidades obtenidas para la totalidad de cuentas que pertenecen a los diferentes segmentos y subsegmentos del modelo. Las barras que se muestran en cada gráfica se refieren a la concentración de cuentas que hay en cada grupo, como se observa en cada una se tiene el 10% de la población de este segmento.

La línea roja muestra la tasa de malos que hay en cada grupo de probabilidad, es decir, cuántas cuentas realmente mostraron 90+ días de impago a los doce meses posteriores de su calificación, del total de cuentas de ese grupo de cuentas. Como se observa en la gráfica 1, en todos los rangos de probabilidad de todos los segmentos y subsegmentos se encuentra monotonidad, es decir, que a mayor rango de probabilidad asignada mayor es la tasa de malos o la cantidad de cuentas que muestran 90+ días de impago a los doce meses posteriores de su calificación de cada rango.

Dado lo anterior, se concluye que las regresiones modeladas son correctas y, consecuentemente, el modelo funciona ya que, se logra el objetivo de identificar aquellas cuentas que podrían presentar un mal compartimento, así como ser un riesgo para el buen desempeño del portafolio de tarjetas de crédito.

Si bien, hasta ahora se ha cumplido con el principal objetivo del modelo de *credit scoring*, el cual es identificar aquellas cuentas que puedan presentar una probabilidad mayor de presentar un impago, el beneficio que se persigue con este tipo de modelos es que sea de fácil comprensión, aplicación y control para el usuario final.

Figura 4
Estabilidad de Segmentos y Subsegmentos



Fuente: Elaboración propia con los resultados de la estimación.

Es por lo anterior, que se busca transformar la probabilidad a un puntaje para su mejor comprensión, es decir, un número discreto de tres dígitos, el cual entre mayor sea se espera que la cuenta presente un buen comportamiento, posterior a su calificación. Esta transformación de la probabilidad obtenida se realiza mediante la aplicación de una sencilla fórmula.

$$Score = Offset - Factor * \ln(odds) \tag{6}$$

Los *odds* o la razón de momios representa la razón entre la probabilidad de cumplir la variable objetivo⁶ y no cumplirla, es decir, que si la probabilidad obtenida mediante las regresiones de una cuenta es del 0.35, su probabilidad de no cumplirla es de $1 - 0.35 = 0.65$, por lo que su razón de momios es de 0.5384.

⁶ 90+ días de impago en 12 meses posteriores a su calificación.

En lo que respecta al elemento *Factor*, como su nombre lo indica es un factor de escalamiento, el cual se obtiene con la siguiente fórmula.

$$Factor = \frac{PDO}{\ln(2)} \quad (7)$$

En donde, PDO significa *points to double* y se refieren a cada cuántos puntos se doblan la razón de momios, *odds*, para el presente modelo será de 20, es decir, cada 20 puntos se dobla la relación de momios (*odds*). Siddiqui (2006) expone que este valor es uno de los más usados en la industria.

En cuanto al componente *Offset*, se elige con base a la siguiente fórmula:

$$Offset = Score - \{Factor * \ln(odds)\} \quad (8)$$

En donde, si se quieren momios de 50:1 es decir, 50 cuentas buenas por una cuenta mala a 713 puntos y como ya se vio, se doblan dichos momios cada 20 puntos (PDO), el offset daría como resultado 600 puntos, como se muestra en el siguiente cálculo.

$$713 - \left\{ \frac{20}{\ln(2)} * \ln(50) \right\} = 600 \quad (9)$$

Continuando con la cuenta donde su probabilidad es 0.35, sustituyendo los valores ya obtenidos y mencionados en la fórmula 1.5 quedaría como se muestra a continuación.

$$Score = 600 - \frac{20}{\ln(2)} * \ln\left(\frac{0.35}{0.65}\right) \quad (10)$$

El resultado de la sustitución es de 617.86, lo que para una mejor administración y aplicación del modelo se redondea a 618.

Como se puede observar, los valores tanto del *Offset* como de *PDO* son elegidos de manera arbitraria, y es decisión de cada desarrollador o institución financiera elegirlos como mejor les convenga. Siddiqui (2006) argumenta que, cuál sea la opción elegida para el escalamiento del modelo, no afecta al poder predictivo del mismo.

Como se argumenta, uno de los grandes beneficios del *credit scoring* es la facilidad con la que una persona, sin necesidad de estar familiarizada con el desarrollo de este, puede aplicarlo y asignar un puntaje a una cuenta, con tan solo conocer el valor de sus atributos. Es por eso por lo que, se elabora una tabla de puntajes, en donde se muestran los puntos que corresponden a cada atributo de cada variable.

Para la construcción de la tabla de puntajes, son necesarios los valores de β , así como del intercepto que se obtuvieron en las regresiones logísticas de cada subsegmento y segmento del modelo. Por otro lado, al igual que en la fórmula anterior se necesitan valores para el *Factor* y el *Offset*.

La fórmula para obtener los puntos que aporta cada atributo de cada variable al puntaje final o score, se obtienen con la siguiente fórmula.

$$- \left(woe_j * \beta_i + \frac{\alpha}{n} \right) * factor + \frac{offset}{n} \quad (11)$$

$$woe = \ln \left(\frac{p_j}{q_j} \right) \quad (12)$$

En dónde:

p = Se refiere a la proporción de malas o que cumplieron la variable objetivo del atributo.

q = Se refiere a la proporción de buenas o que no cumplieron la variable objetivo del atributo.

β_i = Es el valor obtenido de las regresiones de cada variable.

α = Es el intercepto del segmento o subsegmento.

Factor = Factor de escalamiento (véase fórmula 7).

Offset = Puntos de compensación que tendrá una cuenta por el simple hecho de calificarse.

n = Número de variables del segmento o subsegmento.

Como se observa en la fórmula 13, se hace referencia al valor de β por cada variable de cada segmento o subsegmento, sin embargo, como se puntualiza, para el presente trabajo se tiene una β por cada atributo de cada variable. Lo anterior debido a que la β arrojada por la regresión ya contiene el factor de *woe*, por lo que no es necesario multiplicar el valor de β por el *woe*, dado lo anterior la fórmula queda de la siguiente manera:

$$- \left(\beta_j + \frac{\alpha}{n} \right) * factor + \frac{offset}{n} \quad (13)$$

En dónde:

β_j = Es el valor obtenido de las regresiones de cada variable.

α = Es el intercepto del segmento o subsegmento.

Factor = Factor de escalamiento (véase fórmula 7).

Offset = Puntos de compensación que tendrá una cuenta por el simple hecho de calificarse.

n = Número de variables del segmento o subsegmento.

Como se observa en las tablas de resultados de la regresión logística, los últimos rangos o atributos de cada variable no tienen valor, por lo que esta se

calcula sumando los valores de los demás atributos de la variable multiplicándolos por -1.

Aplicando la fórmula 13, para cada uno de los segmentos y subsegmentos se obtienen las tablas de puntaje. En el apéndice 3 se muestran estas para cada segmento y subsegmento del modelo.

Una vez obtenidas las tablas de puntaje de todos los segmentos y subsegmentos del modelo, se identifica a qué segmento pertenece la cuenta a calificar y en qué rangos o atributos se sitúan los valores que contiene la misma en las variables que forman la tabla de puntajes. Por último, se suman los puntos de los rangos en donde se sitúan dichos valores y se redondea para obtener el puntaje o *score* final.

Resultados

Con el objetivo de revisar la estabilidad y la potencia del modelo, se califica mediante las tablas de puntaje los meses del periodo de modelado⁷ y los meses posteriores al mismo (periodo de validación), es decir, se califican todas las cuentas del portafolio de tarjeta de crédito de abril de 2017 a agosto de 2019.

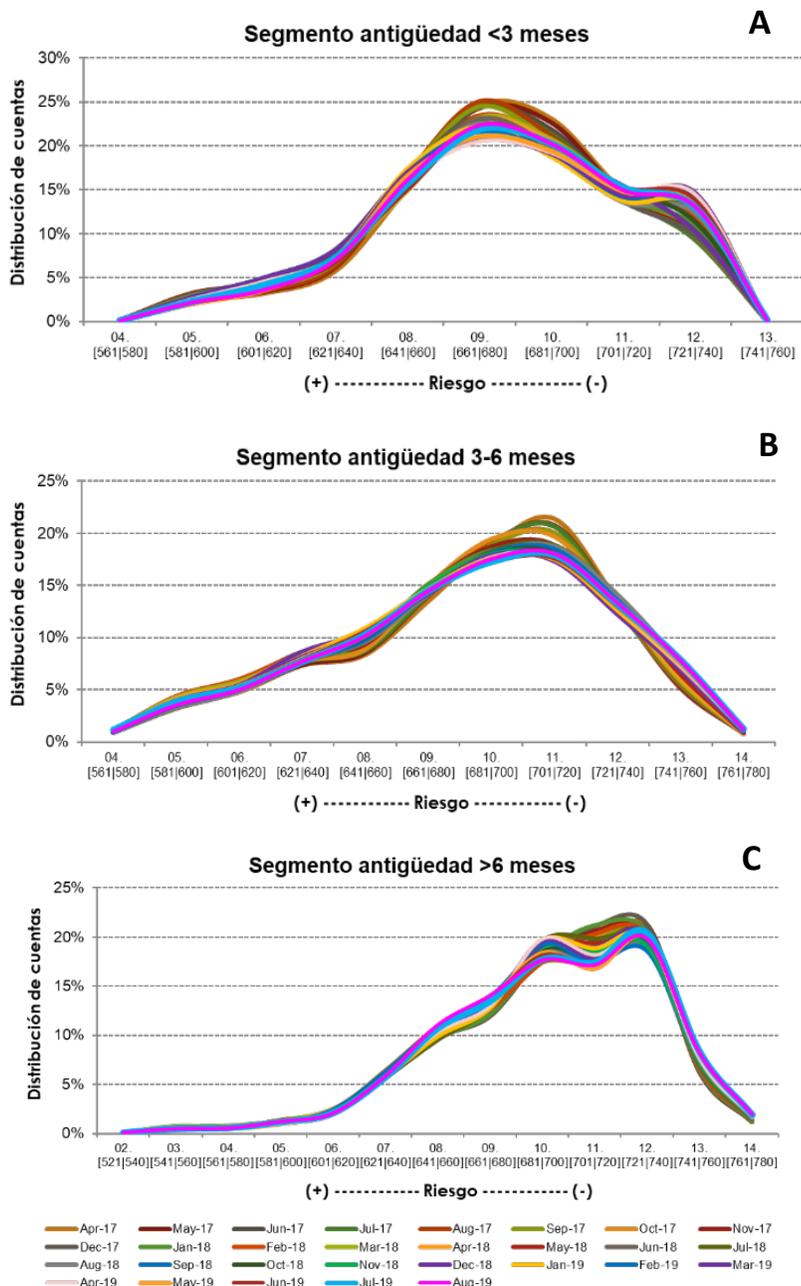
Una vez que cada una de las cuentas del portafolio cuenta con un puntaje o *score*, tanto en el periodo de modelado (abril17 a marzo18) como en el periodo de validación (abril18 a agosto 19), se estiman los rangos equidistantes⁸ de 20 puntos. Dichos rangos permitirán agrupar las cuentas con el objetivo de validar la distribución de estas, mes con mes a través del tiempo. En las siguientes gráficas se muestran los resultados de lo ya mencionado.

Como se observa en las gráficas siguientes, en cada segmento del modelo se presentan 29 líneas, las cuales representan cada mes, desde abril 2017 a agosto 2019, en cuanto al eje de las **X** se muestra la distribución o la densidad de cuentas que se encuentran en cada rango de puntaje o *score* (eje de las **Y**). Como se señala, entre mayor es dicho puntaje menor es la probabilidad de presentar impago, por lo que, se considera que una cuenta con mayor puntaje representa menor riesgo.

⁷ Periodo de abril de 2017 a marzo de 2018

⁸ Que supone encontrarse a la misma distancia.

Figura 5
Distribución de cuentas por rangos de puntaje



Fuente: Elaboración propia con datos de la estimación.

Por otro lado, se observa que, en cada una de las gráficas del grupo denominado gráfica de la figura 5, las densidades o distribuciones de cada segmento del modelo son similares a través del tiempo, esto debido a que todas las líneas correspondientes a cada uno de los meses muestran una tendencia semejante a la de la línea café (Apr-17), la cual corresponde al periodo de modelado. Por lo anterior, se concluye que el modelo, tanto en su totalidad como en sus segmentos, es estable a través del tiempo y asegura una calificación correcta de las cuentas.

Otra forma de validación del modelo, fuera del periodo de modelado, es revisando la monotonicidad de los segmentos de la misma forma como se realiza para el periodo de modelado. Dado que se quiere observar las cuentas doce meses posteriores a su calificación, para comprobar si estas muestran impago, dicha validación se puede realizar de abril de 2017 a agosto de 2018. Los resultados de esta se muestran en las gráficas siguientes:

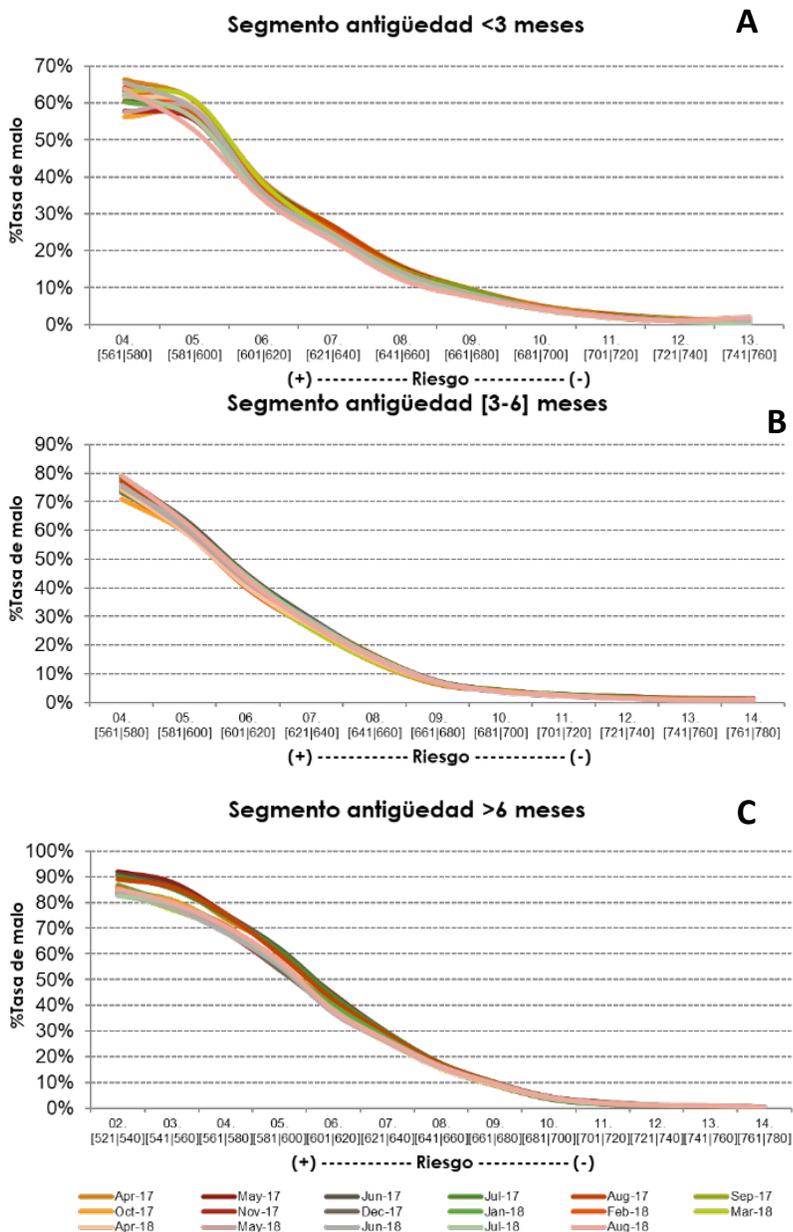
Como se observa en la gráfica de la figura 6, para cada uno de los segmentos se muestra una gráfica en la cual se presenta, por cada rango equidistante de *score* (eje de las Y), el porcentaje de cuentas que cumplieron la variable objetivo⁹ con respecto al total de cuenta que contiene cada rango de *score* (eje de las X). Cada línea se refiere al comportamiento de los rangos durante los meses tanto del periodo de modelado como del periodo de validación.

Dado que la tasa de malos se refiere al porcentaje de malos o de cuentas que incumplieron en los doce meses posteriores a su calificación, se espera que, en los rangos de menor puntaje, esta tasa sea más alta que la de los rangos de mayor puntaje. Como se observa en las gráficas anteriores, todos los segmentos muestran esta monotonicidad y estabilidad en la misma, en los 16 meses posteriores al periodo de modelado (abril-17 a marzo-18), por lo que, se puede concluir que el modelo en su totalidad, y en sus segmentos predice de manera correcta aquellas cuentas que tienen mayor propensión a presentar un impago de 90+ días después de su calificación.

Por último, para asegurar que el modelo guarda la predictibilidad adecuada, se revisa la prueba estadística de KS (Kolmogorov-Smirnov), así como la estabilidad de esta a través del tiempo.

⁹ 90+ días de impago en 12 meses posteriores a su calificación.

Figura 6
Estabilidad tasa de malos periodos de validación



Fuente: Elaboración propia con datos de la estimación.

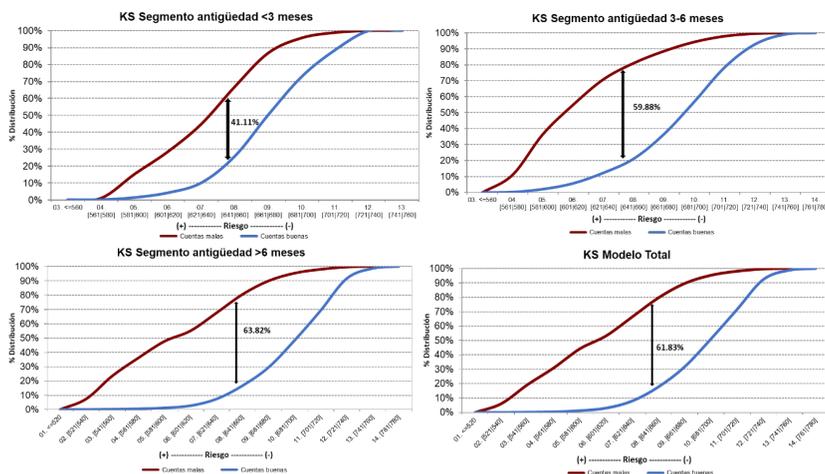
El KS (Kolmogorov–Smirnov) es una prueba no paramétrica, la cual permite conocer la máxima separación existente entre dos funciones de distribución acumuladas, por lo que entre mayor es dicha separación, mayor es el KS y mayor predictibilidad tiene el modelo.

Para el presente modelo, las funciones que se consideran son: las cuentas que se denominan malas, ya que presentan el evento o cumplen la variable objetivo¹⁰, y las cuentas que no presentan el evento o cuentas buenas.

Como se menciona, se realiza una distribución acumulada de dichas funciones por rango de *score*, obteniendo la diferencia entre cada distribución en cada rango, la máxima diferencia obtenida es el valor del estadístico KS.

En las siguientes gráficas, se muestran dichas distribuciones acumuladas por cada segmento del modelo, así como las del modelo en su totalidad, en el periodo de modelado (abril17 a marzo18).

Figura 7
Estadístico KS Modelo total



Fuente: Elaboración propia con datos de la estimación.

Como se observa en la gráfica de la figura 7, para cada segmento del modelo se presenta una gráfica que explica el cálculo del estadístico KS, en esta se muestran las dos distribuciones acumuladas, la línea roja que representa las cuentas malas o que presentaron evento y la línea azul que son aquellas cuentas que no presentaron evento o cuentas buenas. En el eje de las X, se encuentra el

¹⁰ Aquella cuenta que presente 90+ días de impago en 12 meses posteriores a su calificación.

porcentaje de distribución, y en el eje de las ordenadas, los rangos equidistantes¹¹ de score que van de menor a mayor. Una vez observadas las distribuciones acumuladas de las dos funciones, se procede a calcular la máxima separación entre las mismas.

Para el caso del segmento de antigüedad < 3 meses, la máxima separación es de 41.11%, en lo que respecta al segmento de antigüedad 3-6, su KS es de 59.88%, para el segmento de antigüedad >6, su máxima separación es de 63.82%. En cuanto a la totalidad del modelo, esta muestra un KS de 61.83%.

En la siguiente tabla se muestra la interpretación más común en el sistema bancario para los valores del estadístico de KS.

Tabla 16
Referencia valor KS

Valor de KS	Interpretación
<20%	El modelo no es válido
20% - 40%	El modelo es aceptable
41% a 50%	El modelo es bueno
51% a 60%	El modelo es muy bueno
61% a 75%	El modelo es excelente
>75%	El modelo es quizá demasiado bueno para ser verdad

Fuente: Elaboración propia.

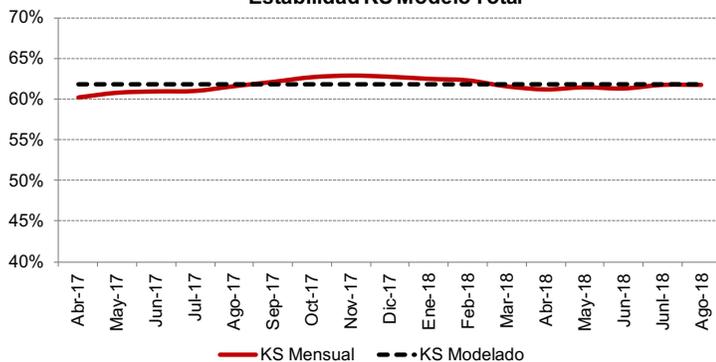
De acuerdo con la tabla 16, los valores obtenidos del estadístico KS para los diferentes segmentos de modelo son buenos y, por consiguiente, el modelo en su totalidad se puede tomar como un modelo excelente.

La última prueba que se realiza al modelo es la estabilidad del estadístico KS, obtenido en el punto anterior, el objetivo es que tanto para los meses del periodo de modelado (de abril 2017 a marzo, de 2018) como en los del periodo de validación (de abril 2018 a agosto, 2018) se mantenga el valor del estadístico KS del modelo total, en 61.83%.

Dado que se pretende observar las cuentas doce meses posteriores a su calificación, para comprobar si estas muestran impago o son cuentas catalogadas como malas, la validación se puede realizar de abril de 2017 a agosto de 2018. En la siguiente gráfica, se presenta el resultado del estadístico KS cada mes del periodo de modelado y de validación.

¹¹ Que supone encontrarse a la misma distancia.

Figura 8
Estabilidad estadístico KS
Estabilidad KS Modelo Total



Fuente: Elaboración propia con datos de la estimación.

Como se aprecia en la gráfica de la figura 8, en el eje de las X se muestran los meses de abril, 2017 a agosto, 2018 y en el eje de las ordenadas, el valor de la máxima separación que existe entre las dos distribuciones (cuentas malas y cuentas buenas). La línea negra punteada se refiere al valor del KS, que se obtiene en el periodo de modelado (61.83%), y la línea roja representa el valor del KS en cada mes.

Como se observa, la línea roja siempre es cercana a la línea negra punteada, por lo que se puede afirmar que el estadístico KS mantiene sus niveles a través de todos los meses del modelado y fuera del mismo.

Complementando las pruebas estadísticas, se realiza la matriz de confusión o de contingencia, la cual se trata de una tabla en donde se clasifica cada cuenta dependiendo del cumplimiento o no de la variable objetivo, tanto con los datos reales como lo predicho por el modelo.

Para el presente modelo, se denominará como “Incumple” a toda aquella cuenta que cumple la variable objetivo, es decir, a toda aquella cuenta que presenta 90 días o más de incumplimiento en los 12 meses posteriores a su calificación. Y a toda cuenta que no presenta dicho incumplimiento (no cumple la variable objetivo), se le denominará “Buena”.

En la tabla 17, se muestran las cuentas modeladas con su clasificación, tomando para el modelo un corte de 660 puntos, es decir, toda aquella cuenta que se la haya asignado mayor puntaje que 660 es considerada como “Buena” y toda cuenta que se le evaluó con un puntaje menor igual que 660 es

considerada como “Incumplen”. Como lo menciona Siddiqui (2006), el corte es establecido por cada institución basado en la tasa de malos de las cuentas. Con base en la matriz de confusión, se pueden obtener diferentes medidas de evaluación, en específico, para los modelos de comportamiento de crédito se busca tener un buen nivel de exactitud y de especificidad.

Tabla 17
Matriz de confusión

		Modelo		
		Buena	Incumplen	Total
Real	Buena	8,846,421	1,894,665	10,741,085
	Incumplen	308,493	754,575	1,063,068
	Total	9,154,914	2,649,239	11,804,153

Fuente: Elaboración propia con datos de la estimación.

La exactitud se refiere al porcentaje de cuentas que el modelo predijo correctamente, es decir, aquellas cuentas que en los datos reales son buenas y el modelo las predijo como tal, y aquellas que en los datos reales incumplen y el modelo acierta en clasificarlas de dicha manera. Por otro lado, la especificidad se refiere a qué tantas cuentas que en los datos reales incumplen y el modelo las predice correctamente. Dichas medidas de evaluación para el presente modelo se calculan de la siguiente forma:

$$\text{Exactitud} = (8,846,421 + 754,575) / 11,804,153 = 81\%$$

$$\text{Especificidad} = (754,575 / 1,063,068) = 71\%$$

Como se muestra, el presente modelo tiene una exactitud del 81%, es decir, el 81% de las cuentas las predice de manera correcta. En lo que respecta a la especificidad, el modelo presenta 71%, es decir, dado que nos interesa evitar Falsos Positivos (cuentas que en los datos reales incumplen, pero el modelo las asigna como buenas), el porcentaje obtenido es bastante bueno.

Para finalizar, las evaluaciones con la matriz de confusión, se obtienen las medidas de exactitud y especificidad para periodo de validación esperando que se mantengan. Para el periodo de validación del presente modelo (de abril de 2018 a agosto de 2018) se obtiene una exactitud del 80% y una especificidad del 70%, por lo que se puede asegurar que el modelo guarda su predictibilidad a través del tiempo.

Dado lo anterior, se puede asegurar que el modelo además de tener una potencia de predicción excelente, esta es estable a lo largo del tiempo, por lo que se asegura una correcta calificación de cada cuenta del portafolio en el futuro próximo.

Conclusiones

La presente investigación cumple el objetivo de diseñar un modelo que identifique a las cuentas del portafolio de tarjeta de crédito y las jerarquice según su riesgo de incumplimiento, mediante un sencillo puntaje. En términos de la hipótesis, se comprueba que el modelo propuesto tiene un alto nivel de predictibilidad y de estabilidad, tanto fuera como dentro del periodo de modelado, aunado con la comprobación de monotonicidad que asegura que el modelo tenga un alto nivel de precisión.

El desarrollo de la investigación integra una serie de pasos agrupados en dos procesos: i) el tratamiento y construcción de datos y ii) la aplicación del modelo, en el cual se incorpora la aplicación de una regresión logística, pruebas de estabilidad del modelo, construcción de tabla de puntajes y validación del modelo.

A partir de los resultados del modelo se conocen las principales variables que determinan la probabilidad de incumplimiento por cada segmento y grupo. Las pruebas de estabilidad del modelo y la validación de este sugieren que la estimación es sumamente confiable para las decisiones orientadas a la mitigación del riesgo de crédito dentro de la institución bancaria. La transformación de la probabilidad a un puntaje facilita la utilización de dicha información, para su uso interno, en la toma de decisiones orientadas al otorgamiento de crédito y oferta de diversos productos ofrecidos por el banco. Por último, se demuestra la estabilidad del modelo en cuanto a su distribución y predicción, mediante una validación fuera del periodo de modelado, lo cual comprueba la viabilidad de uso en el presente y futuro próximo.

La importancia de la investigación subyace en que, además del beneficio para la institución, se beneficia directamente a la población bancarizada, al categorizarla, permite que la institución bancaria tome acciones para el cuidado del endeudamiento de la persona en cuestión, limitando el uso del crédito y evitando su sobreendeudamiento.

Al permear esta práctica a todas las instituciones bancarias, se estaría previniendo el riesgo de crédito a un nivel macroeconómico, por lo que, el sistema financiero en su totalidad gozaría de un comportamiento sano en cuanto al control del riesgo de crédito y, por ende, de los diferentes riesgos del sistema financiero.

Aparte de la mitigación de dichos riesgos, con el desarrollo del modelo presentado, se disminuye el problema de la información asimétrica, ya que al asignar una puntuación al crédito se cuenta con una medición cuantitativa de

las cuentas, basándose en las cualidades de estas, pudiendo numéricamente diferenciar a las mejores cuentas de las peores.

Referencias

- [1] Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., y Bilal, M. (2018). "Systematic review of bankruptcy prediction models: Towards a framework for tool selection." *Expert Systems with Applications*, 94, 164-184. <https://doi.org/10.1016/j.eswa.2017.10.040>
- [2] Ampountolas, A., Nyarko Nde, T., y Constantinescu, C. (2021). "A Machine Learning Approach for Micro-Credit Scoring." *Risks*, 9(3), 50. <https://doi.org/10.3390/risks9030050>
- [3] <https://doi.org/10.3390/risks9030050>
- [4] Ansari, A., Ahmad, I. S., Bakar, A. A., y Yaakub, M. R. (2020). "A hybrid metaheuristic method in training artificial neural network for bankruptcy prediction." *IEEE Access*, 8, 176640-176650. doi: 10.1109/ACCESS.2020.3026529.
- [5] Byström, H. (2019). "Blockchains, real-time accounting, and the future of credit risk modeling". *Ledger*, 4. <https://doi.org/10.5195/ledger.2019.100>
- [6] Capelo Vinza, J. A. (2012). *Modelo de aprobación de tarjetas de crédito en la población ecuatoriana bancarizada a través de una metodología analítica* (Tesis de Licenciatura, Escuela Politécnica Nacional). Recuperado de: <http://bibdigital.epn.edu.ec/handle/15000/5328>
- [7] Caruso, G., Gattone, S. A., Fortuna, F., y Di Battista, T. (2021). "Cluster Analysis for mixed data: An application to credit risk evaluation". *Socio-Economic Planning Sciences*, 73, 100850. <https://doi.org/10.1016/j.seps.2020.100850>
- [8] Cavallo, A., Coghlan, M., & Jackson, J. (2012). *Aplicación de "credit scoring" como solución a los problemas actuales de tarjeta AA en el área de originación y análisis crediticio*. (Tesis de Licenciatura, Instituto Tecnológico de Buenos Aires) Recuperado de: <http://ri.itba.edu.ar/handle/123456789/955>
- [9] Chuang, C. L., y Lin, R. H. (2009). "Constructing a reassigning credit scoring model". *Expert Systems with Applications*, 36(2), 1685-1694. <https://doi.org/10.1016/j.eswa.2007.11.067>
- [10] <https://doi.org/10.1016/j.eswa.2007.11.067>
- [11] Davi, S. S., y Radhika, Y. (2018). "A survey on machine learning and statistical techniques in bankruptcy prediction". *International Journal of Machine Learning and Computing*, 8(2), 133-139. doi: 10.18178/ijmlc.2018.8.2.676
- [12] De la Fuente, S. (2011). *Regresión Logística*. Disponible en: <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/CUALITATIVAS/LOGISTICA/regresion-logistica.pdf>
- [13] Gutierrez Girault, M. A. (2007). "Modelos de credit scoring: qué, cómo, cuándo y para qué." Banco Central de la República Argentina, Argentina. Recuperado de: <https://mpr.a.ub.uni-muenchen.de/id/eprint/16377>
- [14] Ignatius, J., Hatami-Marbini, A., Rahman, A., Dhamotharan, L., & Khoshnevis, P. (2018). "A fuzzy decision support system for credit scoring." *Neural Computing and Applications*, 29(10), 921-937. <https://doi.org/10.1007/s00521-016-2592-1>
- [15] López, P. P. (2018). *Desarrollo de modelo Credit Scoring para admisión de facturas en factoring*. (Tesis de Licenciatura, Universidad de Chile) Recuperado de: <https://repositorio.uchile.cl/handle/2250/149160>

- [16] Medina-Olivares, V., Calabrese, R., Dong, Y., y Shi, B. (2021). "Spatial dependence in microfinance credit default." *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2021.05.009>
- [17] Okesola, O. J., Okokpujie, K. O., Adewale, A. A., John, S. N., & Omoruyi, O. (2017, December). An improved bank credit scoring model: a naïve Bayesian approach. In 2017 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 228-233). IEEE. <https://doi.org/10.1109/CSCI.2017.36>
- [18] Papouškova, M., y Hajek, P. (2019). "Two-stage consumer credit risk modelling using heterogeneous ensemble learning." *Decision support systems*, 118, 33-45. <https://doi.org/10.1016/j.dss.2019.01.002>
- [19] Puertas Medina, R. M., y Martí Selva, M. L. (2013). "Análisis del credit scoring." *RAE-Revista de Administração de Empresas*, 53(3), 303-315. <http://hdl.handle.net/10251/59864>
- [20] Siddiqi, N. (2006). *Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring*. Canada: Simultaneously.
- [21] Siddiqi, N. (2017). *Intelligent credit scoring: Building and implementing better credit risk scorecards*. John Wiley & Sons.
- [22] Torvekar, N., & Game, P. S. (2019). "Predictive analysis of credit score for credit card defaulters." *Int. J. Recent Technol. Eng*, 7(1), 4. https://www.researchgate.net/profile/Pravin-Game/publication/332557433_Predictive_analysis_of_credit_score_for_credit_card_defaulters/links/5f44ec63458515b7294fc74c/Predictive-analysis-of-credit-score-for-credit-card-defaulters.pdf
- [23] Trejo-García, J. C., Ríos-Bolívar, H., y Martínez-García, M. Á. (2016). "Análisis de la administración del riesgo crediticio en México para tarjetas de crédito." *Revista mexicana de economía y finanzas*, 11(1), 103-121. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-53462016000100103&lng=es&tlng=es.
- [24] Vega Mena, M. D. L., Santoyo Cortés, V. H., Muñoz Rodríguez, M., y Altamirano Cárdenas, J. R. (2017). "Reducción de costos de transacción e información asimétrica: experiencias de financiamiento rural en México". *Estudios sociales (Hermosillo, Son.)*, 27(49), 181-209. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-45572017000100181&lng=es&tlng=es
- [25] Zhang, Y., Liu, R., Heidari, A. A., Wang, X., Chen, Y., Wang, M., & Chen, H. (2021). "Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis." *Neurocomputing*, 430, 185-212. <https://doi.org/10.1016/j.neucom.2020.10.038>

Apéndices

Consulte los apéndices [en línea](#).