

Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado

MAESTRÍA EN FINANZAS

TRABAJO FINAL DE MAESTRÍA

Implementación de score crediticio a individuos en
entorno Fintech - periodo 2016-2018

AUTOR: ROBERTO ALFREDO GOMEZ

DIRECTOR: GONZALO RONDINONE

NOVIEMBRE DE 2019

Agradecimientos

A la FCE-UBA, por propiciar el ámbito de encuentro.

Al Dr. Gonzalo Rondinone, por confiar en el potencial de este proyecto y en las cualidades de quien lo escribe.

Al circunstancial jefe y estimado amigo Sergio Ferreira, por la desinteresada asistencia para llegar a buen puerto. En su nombre, al equipo de Riesgos del Banco Columbia por la constante predisposición.

A mis amigos del barrio, por el aguante que excede tiempo y locación. Por ser familia.

A mi familia, por sus buenos deseos.

A mi señora, la patrona, Geor, porque sintetiza todo lo anterior.

À tout le monde.

À tous mes amis.

Resumen

El siguiente trabajo tiene su origen en un fenómeno reciente y disruptivo como ha sido la implementación de nuevas tecnologías en el mundo de las finanzas. Desde los lineamientos para criptomoneda presentados por Satoshi Nakamoto en los albores del siglo XXI hasta nuestros días, los recursos tecnológicos han penetrado en la milenaria cultura financiera para dotarla de herramientas poderosas.

El objetivo de este trabajo es brindar un modelo de *scoring* para créditos personales que incluya variables de carácter digital, para ser utilizado en un entorno de desarrollo *Fintech*, aprovechando las ventajas de un nuevo ecosistema financiero. Esto se ha logrado tras una investigación de diseño mixto, entre el análisis cuantitativo y cualitativo para procesar la información proveniente de entrevistas con actores relevantes del sector.

El resultado alcanzado permite demostrar que proporcionar un e-mail y utilizar el usuario de Twitter asociado mejoran el poder predictivo de un modelo de score crediticio tradicional. Asimismo, permite incluir más sujetos al mercado crediticio al incorporar información barata, accesible y transversal que los modelos de crédito tradicionales actualmente no utiliza.

Se espera que este trabajo de investigación sirva como puntapié para futuros modelos predictivos que no necesiten incluir variables de comportamiento bancario para, de esta forma, contribuir a mejorar el índice de inclusión financiera.

Para futuras investigaciones se sugiere trabajar con un universo de prospectos adquiridos mediante comunicaciones vía WhatsApp o *chat bot* similares, que hayan interactuado con algún aplicativo de forma tal que se pueda realizar minería de texto sobre sus inputs y así mejorar el modelo propuesto.

Palabras clave: Scoring, Fintech, Inclusión Financiera

Índice

| | |
|---|-----------|
| CAPÍTULO I..... | 6 |
| 1.1. INTRODUCCIÓN | 6 |
| 1.2. JUSTIFICACIÓN | 7 |
| 1.3. ESTRUCTURA DEL TFM..... | 9 |
| 1.4. PLANTEAMIENTO DEL TEMA/PROBLEMA..... | 9 |
| 1.5. OBJETIVOS | 11 |
| 1.5.1. <i>Objetivo general</i> | 11 |
| 1.5.2. <i>Objetivos específicos</i> | 11 |
| 1.6. HIPÓTESIS..... | 11 |
| 1.7. METODOLOGÍA..... | 12 |
| CAPITULO II MARCO TEÓRICO..... | 13 |
| 2.1. CONSUMO | 13 |
| 2.2. CRÉDITO Y RIESGO CREDITICIO | 14 |
| 2.2.1. <i>Mora y default</i> | 15 |
| 2.2.2. <i>Pérdida esperada e inesperada</i> | 15 |
| 2.3. MODELOS DE SCORING | 16 |
| 2.3.1. <i>Modelos paramétricos</i> | 17 |
| 2.3.1.1. <i>El modelo logit</i> | 18 |
| 2.3.2. <i>Modelos no paramétricos</i> | 19 |
| 2.4. SISTEMA DE INFORMACIÓN..... | 21 |
| 2.4.1. <i>Métodos de evaluación admitidos</i> | 21 |
| 2.4.2. <i>Clasificación de deudores</i> | 21 |
| 2.4.3. <i>Centrales de información y protección de datos.</i> | 23 |
| 2.5. INCLUSIÓN FINANCIERA..... | 24 |
| 2.5.1. <i>Definición y limitantes</i> | 24 |
| 2.6. ENTORNO FINTECH: MÁS QUE UNA MODA | 25 |
| 2.6.1. <i>Definición y pertinencia</i> | 25 |
| 2.6.2. <i>Redes sociales</i> | 26 |
| 2.6.3. <i>Mailing</i> | 27 |
| CAPÍTULO III MODELIZACIÓN TRADICIONAL | 28 |
| 3.1. UNIVERSO ANALIZADO..... | 28 |
| 3.1.1. <i>Variables sociodemográficas</i> | 28 |
| 3.1.2. <i>Variables de producto</i> | 30 |
| 3.1.3. <i>Variables de comportamiento crediticio</i> | 31 |
| 3.1.4. <i>Variable objetivo</i> | 31 |
| 3.2. ANÁLISIS BIVARIADO: | 31 |
| 3.2.1. <i>Variables sociodemográficas</i> | 32 |
| 3.2.2. <i>Variables de producto</i> | 32 |
| 3.2.3. <i>Variables de comportamiento crediticio</i> | 33 |
| 3.3. MODELO TRADICIONAL..... | 34 |
| 3.3.1. <i>Variables del modelo</i> | 35 |
| 3.4. PRUEBAS DE ESTIMADOR..... | 38 |
| 3.4.1. <i>Matriz de clasificación</i> | 38 |
| 3.4.2. <i>Scorecard</i> | 39 |
| 3.4.3. <i>Cálculo de Hosmer – Lemeshow:</i> | 40 |
| 3.4.4. <i>Curva ROC y área bajo la curva (AUC)</i> | 41 |
| 3.5. TESTEO..... | 42 |
| CAPÍTULO IV MODELIZACIÓN DIGITAL | 45 |
| 4.1. VARIABLES DIGITALES INCORPORADAS..... | 45 |
| 4.1.1. <i>Variables asociadas al e-mail</i> | 45 |
| 4.1.2. <i>Variables asociadas al número telefónico</i> | 46 |
| 4.1.3. <i>Variables asociadas a la cuenta de Twitter</i> | 47 |

| | | |
|---|--|-----------|
| 4.2. | ANÁLISIS BIVARIADO | 47 |
| 4.3. | MODELO DIGITAL | 48 |
| 4.4. | PRUEBAS DE ESTIMADOR..... | 51 |
| 4.4.1. | <i>Matriz de clasificación</i> | 51 |
| 4.4.2. | <i>Scorecard</i> | 52 |
| 4.4.3. | <i>Cálculo de Hosmer – Lemeshow:</i> | 53 |
| 4.4.4. | <i>Curva ROC y área bajo la curva (AUC)</i> | 53 |
| 4.5. | TESTEO..... | 54 |
| 4.6. | CONSIDERACIONES | 55 |
| CAPÍTULO V CONCLUSIONES | | 57 |
| REFERENCIAS BIBLIOGRÁFICAS | | 60 |
| ANEXO I | | 65 |
| ANEXO II | | 66 |
| ANEXO III | | 67 |
| | ANÁLISIS BIVARIADO DE LAS VARIABLES CONSIDERADAS PARA REALIZAR MODELO TRADICIONAL..... | 70 |
| ANEXO IV | | 81 |
| | ANÁLISIS BIVARIADO DE LAS VARIABLES CONSIDERADAS PARA REALIZAR MODELO DIGITAL. | 82 |

Capítulo I

1.1. Introducción

Una frase muy común utilizada en el mundo financiero ayuda a dar pie a la discusión “Un banquero es una persona dispuesta a prestarte dinero si puedes probar que no lo necesitas”. Esta proposición, atribuida a Herbert V. Prochnow¹, pone de manifiesto una de las mayores paradojas inherentes a la actividad crediticia, como lo es el buscar en el beneficiario del préstamo una solvencia tal que haga improbable que necesite dinero prestado.

Esta posición conservadora está influenciada por la dificultad de una estimación certera en torno a la real posibilidad de repago del prestatario. Resulta extremadamente complejo estimar el comportamiento futuro de un cliente, predecir no solo su capacidad de pago sino su voluntad de honrar sus compromisos. Entonces, en un escenario de riesgo conservador, se castiga a todo el universo de potenciales clientes con condiciones de acceso al crédito muy restrictivas o directamente se lo excluye. La revolución tecnológica contemporánea y la posibilidad de acceso a grandes volúmenes de información para ser analizada pueden aportar soluciones a esta problemática.

Los datos que los propios usuarios suministran en distintas plataformas y redes sociales permiten tener un mayor conocimiento de los potenciales tomadores de créditos. En este contexto, cabe preguntarse si la inclusión de esta información en modelos tradicionales permite mejorar la toma de decisiones, mitigar el riesgo, y por sobre todas las cosas, incorporar la mayor cantidad de ciudadanos al sector financiero formal.

Se realizará en este trabajo un repaso de la literatura tradicional de modelos de *scoring* que se utilizan en la banca, comenzando con el emblemático *paper* de Loretta J. Mester (1997) que explica las ventajas de desarrollar *scores* para evaluar riesgos de crédito en préstamos, basado asimismo en el previo trabajo de Edward Altman (1968), pionero en la predicción de quiebras basadas en el comportamiento previo del sujeto de crédito. Se postulará aquí un caso empírico de *scoring* tradicional mediante la elección de la regresión logística binaria por sobre otro tipo de regresiones para predecir eventos dicotómicos, como el pago/*default*. Para dicho caso se utilizarán las variables tradicionales utilizadas, entre otros, por el Banco Columbia y funcionará como *benchmark* de comparación. Se propondrá,

¹ “A banker is a person who is willing to make you a loan if you present sufficient evidence to show you don't need it.”

a modo de aporte, la inclusión de variables “digitales” que se consideran relevantes para la evaluación crediticia online para el segmento de consumo. Finalmente se presentan las conclusiones de este trabajo.

El objetivo es aportar evidencia que un modelo de scoring para créditos personales que incluya variables de carácter digital para ser utilizado en un entorno *Fintech* tiene mayor poder predictivo que uno que solo cuenta con variables tradicionales.

Este trabajo focaliza su análisis en el periodo 2016-2018, dado el cambio de Gobierno Nacional de 2015 y la consecuente renovación de autoridades en el Banco Central de la República Argentina (BCRA), que implicó una gestión más proclive a la implementación de herramientas tecnológicas financieras. Además, se apuntala en el acceso a la base de clientes de créditos minoristas del Banco Columbia, necesaria para validar el modelo de scoring, así como en estadísticas oficiales del Banco Central para el periodo de estudio, que permiten incorporar dimensiones relacionadas con comportamiento crediticio.

1.2. Justificación

Según un relevamiento del Banco Mundial, solo el 49% de los argentinos tiene una cuenta bancaria o relación con una entidad financiera, sin importar el sexo o si se encuentran ocupados. (Demirgüç-Kunt, Klapper, Singer, Saniya, y Hess, 2018). Este dato, por contraposición, arroja que más de la mitad de la población se encuentra fuera del sistema financiero. Implica que un 51% del total que no consigue armar un historial crediticio, lo que restringe su posibilidad de acceder a condiciones igualitarias de financiación, o bien a no acceder a ninguna.

Se puede citar a modo de ejemplo extremo a la fecha de redacción de este trabajo la empresa Préstamo Móvil S.A. ofrecía, a través de su marca “vivus.com.ar” un préstamo de no más de \$7.000² a devolver a 30 días, con un costo financiero total de \$1.398, lo que equivale a un 19% a 30 días, o un 806% anual. Si se compara con el costo de adquisición del dinero cuya tasa de referencia a la fecha de cálculo ascendía a 36,31% anual³ la diferencia luce difícil de explicar, pero termina siendo de los pocos vehículos que aquellos sin historial crediticio pueden acceder.

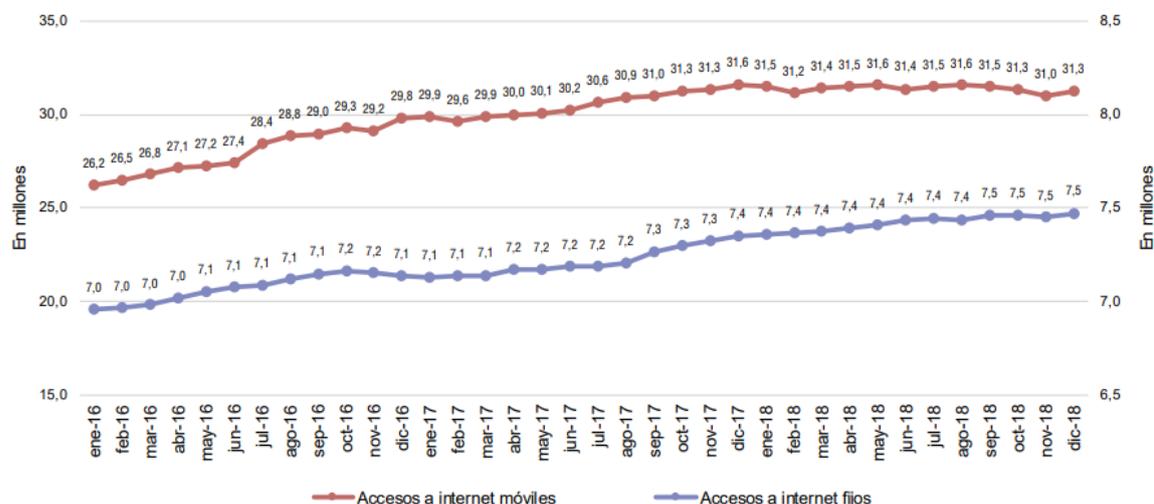
Al mismo tiempo, se observa una tendencia creciente a la conectividad a Internet, tanto en el hogar como en forma inalámbrica (INDEC, 2018). Para el trienio 2016-2018, Argentina superaba las 38 millones de conexiones, como se observa en la Figura 1. En

² Equivalentes a 111 dólares al tipo de cambio oficial el 16/01/2020

³ Tasa BADLAR al 16/01/2020

simultáneo, la penetración de la telefonía celular sobre la población⁴ en este período fue del 150% (Sticco, 2018).

Figura 1 – Accesos a internet fijos y móviles. Enero 2016-diciembre 2018



Fuente: INDEC

En este entorno, las redes sociales han cobrado alta preponderancia. Un reporte de la consultora Kepios indica que Argentina cuenta con 34 millones de usuarios de redes sociales, de los cuales 31 millones lo hacen desde un smartphone, dispositivo con el que cuenta el 73% de la población (Kepios, 2019). Como se mencionó anteriormente esto genera una fuente adicional de grandes volúmenes de información que puede ser utilizada por distintos *stakeholders*.

Asimismo, en el ámbito gubernamental, el cambio de autoridades en el Banco Central, alineado con el cambio de signo político en el Poder Ejecutivo acaecido en 2015, funcionó como punto de partida para la implementación de herramientas tecnológicas en el ecosistema financiero local. La meta: mejorar los estándares de inclusión financiera, alineados a la política rectora del G20⁵, que Argentina presidió durante 2018 (G20, 2018).

Este contexto propicia, justifica y estimula la investigación académica para incorporar herramientas del mundo digital que ayuden a mejorar estrategias y modelos de *scoring* de evaluación crediticia para personas no bancarizadas.

⁴ Penetración de la telefonía celular = Cantidad de suscripciones al servicio de telefonía celular / cantidad de habitantes.

⁵ G20, o “Grupo de los 20”, es un encuentro interanual de Bancos Centrales y Ministerios de Economía de 20 estados emergentes y centrales, tendiente a lograr un crecimiento económico global estable y sostenible en beneficio de todos los países. Más información en <https://www.g20.org/en/summit/about/>

1.3. Estructura del TFM

A continuación, se describe brevemente el ordenamiento del trabajo.

En el primer bloque se desarrolla tanto la **Introducción** como el **Enfoque Metodológico**, en el que se pretende realizar la presentación y descripción del tema, así como su relevancia. Incluye el planteamiento de la temática, y objetivos, tanto general como específicos, además de la Hipótesis rectora y la metodología utilizada para alcanzar las metas propuestas.

El bloque 2 abarca en su totalidad el **Marco Teórico**, entendido como el estado actual de conocimiento y punto de partida para este trabajo. Se presentan los principales conceptos y herramientas necesarias para la confección de un modelo de scoring, así como una contextualización del entorno *FinTech* y las posibles herramientas a utilizar.

Al llegar al bloque 3 se desarrolla la **Modelización Tradicional**, entendida como la presentación de la forma actual para estimar el comportamiento crediticio, con la identificación de variables que se consideran relevantes para la evaluación crediticia online en el segmento de créditos de consumo. Se incluye en este paso una descripción del set de datos recabados, incluyendo cantidad, procedimiento de obtención y caracterización de las principales variables.

El bloque 4 incorpora variables digitales al modelo tradicional para transformarlo en una herramienta de **Modelización Digital**. Para esto se identifican acciones de recolección de datos digitales que sean aplicables al mercado crediticio de consumo. A posteriori, se testean variables de e-mail y social media que puedan ser incorporadas a futuros modelos de scoring crediticio.

El trabajo concluye con un bloque destinado a **Conclusiones**, en las que se describirán las implicancias del modelo, un resumen de las principales ideas propuestas, así como recomendaciones para continuar y profundizar esta línea de investigación.

1.4. Planteamiento del tema/problema

De acuerdo con el Banco Central de la República Argentina (BCRA), la inclusión financiera se entiende como “el acceso y el uso de una oferta amplia de servicios financieros, provistos de una manera sostenible y responsable” (2018). La entidad dedica gran parte de sus acciones a lograr este objetivo. Así, las medidas regulatorias que apuntan a la reducción de costos de las entidades –en particular la digitalización de algunas operaciones- tienen un fin último de inclusión. Esto se sustenta en la idea de que menores costos tienen el potencial

del desarrollo de modelos de negocios que permitan atender a atender segmentos de menores ingresos de la población. Al mismo tiempo la inclusión financiera fue un componente clave en la agenda de la presidencia argentina del G20 durante el año 2018 y en este sentido se encolumnaron las acciones gubernamentales antes descriptas (G20, 2018).

Asimismo, la agenda de Medios de Pago Electrónicos impulsada por el BCRA es un pilar fundamental para lograr una mayor inclusión financiera. La ampliación de la red de aceptación de estos pagos es clave para potenciar su uso por quienes se encuentran bancarizados de manera precaria, es decir, sin hacer uso de sus cuentas bancarias o de sus medios de pago asociados (2018). Todo esto se desarrolla en un contexto de crecimiento del ecosistema *Fintech* en el país. Para 2018 se contabilizaban 133 empresas y, de las 52 firmas consultadas, 18 habían sido constituidas entre 2016 y 2017 (Accenture, 2018, p.13).

A modo enunciativo de los avances en este campo se destaca el pago electrónico (Pago Fácil), el *crowdfunding*⁶(Idea.me) y los puntos de venta móviles (Mi POSNET), que están instalados y forman parte del ecosistema económico local. La presentación de TransferWise presupone un avance en servicios P2P⁷. También el Banco de la Nación se alinea a este nuevo paradigma, al propiciar una política de incentivo al envío de dinero y pagos a través del celular con su plataforma PIM. Todo esto configura un avance en el desarrollo de información digital que puede ser aprovechada por los modelos tradicionales.

Sin embargo, este nuevo ecosistema no ha logrado rebatir una de las principales barreras de acceso sistémico al mercado crediticio, como lo es la “carpeta” de historial de crédito. Para evaluar la capacidad de repago de una persona y decidir sobre el otorgamiento de una línea de financiamiento, las entidades ponderan el comportamiento de sus potenciales clientes en el sistema financiero a través de bureaus de información crediticia. Si la persona nunca recibió asistencia, se la considera “población débil”, lo que reduce sus posibilidades de acceder a un crédito y, consecuentemente, de poder postularse a una nueva asistencia. De esta forma, quedan excluidos del canal formal y deben recurrir a entidades que compensan la falta de información sobre el cliente con sobretasas de interés.

De allí se deriva otra de las barreras de ingreso, dada por las altas tasas de interés que prejuzgan y castigan con sobrepuestos determinados riesgos que no pueden medir con facilidad. Para el 21 de diciembre de 2018, el BCRA informaba una TNA del 82,08% (BCRA, 2019) mientras que la inflación de 2018 fue de 47,6% (INDEC, 2019).

⁶ *Crowdfunding*: Financiamiento colectivo

⁷ P2P o *peer to peer* refiere a tecnología que une punta a punta oferentes con demandantes, sin intermediarios.

Aquí es donde la tecnología podría ayudar, al indizar información a priori cualitativa, como puede ser la reputación on-line de un individuo o si posee una cuenta de e-mail asociada a su nombre, para agregarla a una medición de riesgo tradicional. La pregunta rectora de este trabajo será entonces: ¿Qué características de e-mail y social media pueden agregarse a la calificación crediticia para mejorar la inclusión de los argentinos en el sistema financiero local?

Adicionalmente se planteará: ¿cuáles son las variables que se consideran relevantes para la evaluación crediticia en el segmento de créditos personales?, ¿qué variables se pueden incorporar a futuros modelos de scoring crediticio? y, finalmente, ¿qué recursos Fintech pueden ser aplicables al mercado crediticio de consumo en Argentina?

1.5. Objetivos

1.5.1. Objetivo general

Proponer un modelo de evaluación crediticia que sea aplicable al mercado de créditos de consumo e incluya variables digitales de e-mail y social media, para ser utilizados en un entorno Fintech, considerando el período 2016-2018.

1.5.2. Objetivos específicos

- o Identificar y caracterizar variables que se consideren actualmente relevantes para la evaluación crediticia online en el segmento de créditos de consumo.
- o Testear variables de e-mail y social media que puedan ser incorporadas a futuros modelos de scoring crediticio.
- o Identificar herramientas de recolección de datos digitales que sean aplicables al mercado crediticio de consumo.

1.6. Hipótesis

La incorporación de variables digitales como “zona geográfica”, “e-mail a su nombre”, “perfil social actualizado” o “cantidad de seguidores” mejora los modelos tradicionales y predice de forma más eficiente la probabilidad de pago de créditos de consumo permitiendo incorporar nuevos tomadores de créditos que actualmente se encuentran excluidos por falta de información.

1.7. Metodología

El estudio se realizará con un enfoque mixto, combinando componentes cualitativos y cuantitativos en el mismo proyecto. Se fundamenta esta elección en el pragmatismo, orientación entendida como la búsqueda de alternativas prácticas y alcanzables para plantear, contextualizar e interpretar conceptos (Hernández Sampieri, Fernández Collado, Baptista Lucio, 2010).

Para analizar cuáles son las variables relevantes para la evaluación crediticia en el segmento de consumo se recolectará información de entrevistas a expertos del Banco Columbia. Del mismo modo para indagar variables sociales que puedan ser incorporadas a futuros modelos de *scoring* crediticio y para identificar herramientas utilizadas en el entorno Fintech que sean aplicables al mercado crediticio de consumo.

El planteamiento de las hipótesis será de tipo correlacional, buscando relación entre la visibilidad online y la falta de pago de obligaciones crediticias (Hernández Sampieri, et al., 2010, p81). Se describen estos puntos en función de cada objetivo específico en la tabla síntesis incorporada como anexo I.

Capítulo II Marco teórico

2.1. Consumo

El consumo, en tanto acción y efecto de consumir, se entiende como la utilización de bienes para satisfacer necesidades o deseos de cada individuo (RAE, 2018). Ahora bien, ¿qué tanto es posible alcanzar estas pretensiones? ¿Existe una relación entre el ingreso disponible y la acción de consumir?

Para responder estas preguntas, es necesario citar a John Maynard Keynes, quien definió en 1936 que la “Propensión a Consumir” consiste en la “relación funcional entre Y , un nivel de ingreso dado, medido en unidades de salario, y C , el gasto que para el consumo se toma de dicho nivel de ingreso” (Keynes, 1965, p.88). De esta función, el economista británico concluye que

La suma que la comunidad gasta en consumo depende evidentemente, de 1) el monto de su ingreso, 2) otras circunstancias objetivas que lo acompañan, y 3) las necesidades subjetivas y las inclinaciones psicológicas y hábitos de los individuos, así como de los principios según los cuales se divide el ingreso entre ellos (Keynes, 1965, p.88).

Durante muchos años estas premisas sirvieron como base para que diversos autores mostraran teorías empíricas superadoras. Simon Kuznets, Arthur Smithies, James Duesenberry y Franco Modigliani expusieron correcciones a los principios keynesianos de la función de consumo (Morettini, 2002, p.12). Milton Friedman, por su parte, propuso en 1957 individualizar al consumo en dos factores: el permanente (“*permanent consumption*”) y el transitorio (“*transitory consumption*”). El primero, de tipo permanente, se espera inmutable en el futuro y el segundo, llamado temporario, se considera no se mantendrá en el tiempo (Friedman, 1970).

La hipótesis de Friedman se basa en la premisa de que el consumo depende mayormente del ingreso permanente. Aun cuando los consumidores también gasten parte de su renta temporaria lo hacen en una proporción mucho menor, ya que ahorran la gran parte de ella. Consecuentemente se puede exponer al consumo como función del ingreso permanente con la igualdad $C = cYP$

Friedman afirma entonces que la fórmula de Keynes no era correcta, al suponer que el consumo depende del ingreso actual, y que en ese error se encontraba la contradicción con los estudios empíricos. La nueva hipótesis demostró empíricamente que las familias con

mayor porcentaje de ingreso temporario deben registrar niveles de consumo menores que los hogares que poseen mayor proporción del ingreso de carácter permanente (Morettini, 2002).

2.2. Crédito y riesgo crediticio

Dado que el ingreso disponible no oficia como limitante para la acción de consumo, aparece en escena el crédito, del latín *creditum* -lo que es confiado (Oxford latín dictionary, 1968) - entendido como la “cantidad de dinero u otro medio de pago que una persona o entidad, especialmente bancaria, presta a otra bajo determinadas condiciones de devolución” (RAE, 2018), y orientada a la consecución de diversos fines.

Los créditos siempre han sido otorgados a los diferentes actores de la sociedad para adquirir diferentes productos: a los campesinos para la adquisición de aperos de labranza y pagar sus deudas; a los industriales para incrementar su capacidad de fabricación; a los comerciantes para la adquisición de productos y equipos de distribución que faciliten sus actividades; a los señores feudales con el fin de adquirir armamento y contratar a los soldados necesarios para lidiar con otros feudos; y a los gobiernos, quienes necesitan saldar sus deudas (Morales Castro y Morales Castro, 2014. p.2).

Esta transacción asume implícitamente que los deudores devuelven el capital que les fue cedido junto con la adición de los intereses pactados. Esta asunción puede revelarse como incorrecta dado que concurre una probabilidad de incumplimiento a la que se llamará “riesgo de crédito”.

Existe un relativo consenso a la hora de definir este concepto. El riesgo de crédito se define como la posibilidad de caer en una pérdida económica derivada de la imposibilidad de la contraparte para cumplir con su compromiso de pago (Jorion, 2007, p.430). Esta aseveración, si bien sintética, explica claramente las consecuencias para el prestamista (pérdida económica), identifica al prestatario (la contraparte) y cierra el riesgo al cumplimiento de la obligación de devolver el dinero.

Previamente, en su publicación del año 2000, el comité de Basilea había definido al riesgo de crédito en un sentido similar, pero incorporando el concepto de manejo del riesgo de crédito. El foco estaría puesto en maximizar la tasa ajustada de retorno manteniendo la exposición a riesgo dentro de parámetros aceptables, tanto para la cartera de créditos como para cada uno de los préstamos que la componen. La organización recomendaba desde ese

momento un efectivo manejo del riesgo de crédito, al postularlo como un componente esencial para el éxito a largo plazo de cualquier entidad bancaria (p.1).

En similar sentido se expresa el francés Joël Bessis (2002, p.13) al postular que el riesgo de crédito es el más importante que afronta una entidad bancaria y asocia la definición de riesgo de crédito con la de riesgo de *default*, entendido como el riesgo a que un cliente no pague, y desencadene una pérdida total o parcial del dinero prestado al cliente.

2.2.1. *Mora y default*

Ahora bien, no todo atraso o incumplimiento de las condiciones originarias se considera *default*. Un cliente ingresa en mora (*past due*) cuando mantiene una exposición de cualquier monto relativo al préstamo (capital, intereses o gastos) que no haya sido pagada en su totalidad a la fecha de vencimiento pactado (Comité de Basilea, 2016, p.9). Para que esta exposición no saldada sea considerada *default*, deberán seguirse los lineamientos incluidos en el párrafo 452 del *framework* de Basilea II

Se considera que el incumplimiento con respecto a un deudor en concreto ocurre cuando acontece al menos una de las siguientes circunstancias:

- El banco considera probable que el deudor no abone la totalidad de sus obligaciones crediticias frente al grupo bancario, sin recurso por parte del banco a acciones tales como la realización de protecciones (si existieran).
- El deudor se encuentra en situación de mora durante más de 90 días con respecto a cualquier obligación crediticia importante frente al grupo bancario. Se considerará que los descubiertos se encuentran en situación de mora cuando el cliente haya excedido un límite recomendado o cuando se le haya recomendado un límite inferior al actual saldo deudor (Comité de Basilea, 2004, p.86).

2.2.2. *Pérdida esperada e inesperada*

Los riesgos antes definidos deben poder medirse de alguna manera, de forma tal de poder incorporar ese costo al negocio y asegurar su sustentabilidad en el largo plazo. Para esto existen dos indicadores generalmente aceptados: El cálculo de pérdida esperada (PE) y del capital económico (CE).

La pérdida esperada es un ratio que multiplica la probabilidad de *default* de la cartera, la tasa de recupero de esos créditos y la exposición hacia los mismos. Refleja entonces el valor promedio de las pérdidas y, al considerarse costo del negocio, se la asocia a la política

de provisiones de la entidad financiera. El capital económico, por otro lado, es aquel que se considera necesario para cubrir las pérdidas inesperadas, basadas en la posibilidad de que las pérdidas efectivamente acaecidas sean superiores a las esperadas, por lo que se cubren con capital (BBVA, 2012).

Dentro de los componentes de la pérdida esperada se encuentran lo siguiente:

- 1) PD, o probabilidad de *default* de un préstamo en un horizonte de un año.
- 2) LGD, *loss given default*, que corresponde a la tasa de exposición y se calcula como: 1 - tasa de recupero.
- 3) EAD, o el monto de la cartera expuesto a *default*.
- 4) M, o *maturity*.

Para una *maturity* determinada, se puede calcular la pérdida esperada (EL, *expected loss*) con la siguiente fórmula:

$$EL = PD * LGD * EAD$$

Con esta igualdad se puede determinar el monto esperado a perder en el plazo M tal como se muestra en Schuermann (2004).

2.3. Modelos de scoring

De la ecuación planteada en 2.2.2 se visualiza la importancia de poder calcular la PD para determinar así las potenciales pérdidas del negocio y poder preverlas. Es en este espacio donde los modelos de *scoring* cobran relevancia. Son entendidos como un método de evaluar el riesgo crediticio de solicitudes de préstamos, utilizando datos históricos y técnicas estadísticas para aislar los efectos de determinadas variables en la probabilidad de que un individuo entre en *default*. Estos modelos devuelven una puntuación o "*score*" que el banco puede utilizar no solo para calificar al cliente, sino también para clasificar a sus deudores en forma relativa.

Un modelo bien designado debería arrojar un alto porcentaje de aciertos al predecir si un beneficiario podrá o no hacer frente a sus compromisos. Sin embargo, no existe el modelo perfecto y algunos deudores bien calificados podrían ser en realidad malos pagadores, y viceversa (Mester, 1997).

Para diseñar un modelo correctamente calibrado resulta imprescindible estudiar las relaciones existentes entre la información recogida de cada uno de los créditos concedidos en el pasado y los impagos observados. Una vez realizado este análisis, y utilizando un sistema de puntuación establecido en función de las características del cliente, se podrá determinar su probabilidad de *default*.

Esta PD puede especificarse mediante la siguiente expresión:

$$PD = f(X_1, \dots, X_n) + e$$

Donde:

X_i serán los atributos del sujeto;

e , la perturbación aleatoria;

$f(x)$, la función que determina la relación existente entre las variables utilizadas y PD la probabilidad de que el crédito resulte fallido.

El objetivo principal de estos modelos se basa en estimar la función que permita ajustar con la máxima exactitud las observaciones de la muestra, de manera que el error incurrido en la predicción de incumplimiento sea mínimo. Dependiendo de que la forma de la función, $f(x)$, sea conocida o desconocida se tratará de modelos paramétricos o no paramétricos (Puertas Medina, R. y Martí Selva, M., 2013). En los siguientes apartados se expondrán las características distintivas de ambos modelos, así como sus principales exponentes.

2.3.1. Modelos paramétricos

En un modelo paramétrico la forma de la función $f(x)$ es conocida, por lo que la principal dificultad radica en determinar los parámetros que la componen. Sin embargo, la relación que existe entre la variable independiente (PD) y los parámetros que la integran son diferentes, dependiendo del método estadístico que se utilice, ya sea un modelo de probabilidad lineal, logit, probit o de análisis discriminante (Puertas Medina, R. y Martí Selva, M., 2013).

Los tres primeros se ubican dentro de las técnicas estadísticas estándar para estimar la probabilidad de incumplimiento basadas en datos históricos sobre el desempeño del préstamo y las características del prestatario. El modelo de probabilidad lineal asume que existe una relación lineal entre la probabilidad de incumplimiento y los factores; el modelo logit asume que la probabilidad de incumplimiento está distribuida logísticamente; y el

modelo probit supone que la probabilidad de incumplimiento tiene una distribución normal. El análisis discriminante, por su parte, difiere en que, en lugar de estimar la probabilidad de incumplimiento de un prestatario, divide a los prestatarios en clases de alto y bajo riesgo de default (Mester, 1997).

2.3.1.1. El modelo logit

Este modelo, también llamado de regresión logística binaria, establece la relación entre la probabilidad de que ocurra un suceso determinado para un conjunto de valores observados, por ejemplo, el ingreso de una persona. Siguiendo a Gujarati y Porter (2010), esta probabilidad, se expresa como:

$$Pi = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}}$$

Esta ecuación se puede reinterpretar de la siguiente manera, en lo que se conoce como función de distribución logística acumulativa:

$$Pi = \frac{1}{1 + e^{-(Z_i)}} = \frac{e^Z}{1 + e^Z}$$

con $Z_i = \beta_1 + \beta_2 X_i$

De la ecuación precedente, y siguiendo con lo expresado por Gujarati y Porter (2010), se puede comprobar que, para cada valor de Z_i en el rango $\pm \infty$, Pi toma valores que van de 0 a 1, así como Pi no está linealmente relacionado con Z_i ergo, con X_i .

Considerando entonces a Pi como la probabilidad de ocurrencia, se puede afirmar que $(1 - Pi)$ es la probabilidad de no ocurrencia de este evento binario que se trata de predecir.

$$(1 - Pi) = \frac{1}{1 + e^{(Z_i)}}$$

Por consiguiente, se concluye que:

$$\frac{Pi}{1 - Pi} = \frac{1 + e^{(Z_i)}}{1 + e^{-(Z_i)}} = e^{(Z_i)}$$

Dado que $\frac{P_i}{1-P_i}$ es la razón de las probabilidades en favor de que un evento suceda respecto de la probabilidad de que no suceda, si $P_i = 0.8$, significa que las posibilidades son 4 a 1 en favor de que el evento suceda (Gujarati, D. y Porter, D., 2010, p.554).

Aplicando el logaritmo natural de la última ecuación, obtenemos un resultado muy interesante, a saber,

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_1 + \beta_2 X_i$$

es decir, L, el logaritmo de la razón de las probabilidades, no es sólo lineal en X, sino también (desde el punto de vista de estimación) lineal en los parámetros. L se llama logit, y de aquí el nombre modelo logit para modelos como el aquí descrito (Gujarati, D. y Porter, D., 2010, p.554).

2.3.2. Modelos no paramétricos

Así como los modelos paramétricos parten de la premisa de contar con una función $f(x)$ conocida, los modelos no paramétricos emplean funciones estimativas, de la forma $\hat{f}(x)$ que permiten formular una función que sea una buena aproximación de $f(x)$. Los principales exponentes de esta clase de modelos son árboles de decisión.

Los árboles de decisión son una técnica no paramétrica que reúne las características del modelo clásico univariante y las propias de los sistemas multivariantes.

Originariamente fueron propuestos para separar las observaciones que componen la muestra asignándolas a grupos establecidos a priori, de forma que se minimizara el coste esperado de los errores cometidos (Puertas Medina, R. y Martí Selva, M., 2013).

Los árboles de decisión trabajan con clasificadores, características observables, que permiten armar un conjunto de nodos. Los clasificadores se construyen a través de datos históricos recabados a tal efecto, llamados “muestras de aprendizaje” (Breiman, Friedman, Olsen, Stone, 1984).

En términos matemáticos, el modelo CART, acrónimo de *Classification And Regression Tree*, trabaja con el supuesto de que las observaciones son extraídas de una distribución ϕ en $L \times X$, donde L es el espacio de categorías, y X el espacio de características. Las densidades condicionales $\phi(x|l)$ difieren al variar l , y las probabilidades marginales

$\phi(l)$ son conocidas. El proceso utiliza la muestra S como un subconjunto de entrenamiento para la estimación no paramétrica de una regla de clasificación que permita particionar directamente el espacio X de características. Para cada l de L , el subconjunto S_l del conjunto de entrenamiento S constituye una muestra aleatoria de la distribución condicional $\phi(x|l)$ en X .

Este procedimiento se basa en dividir la muestra original en diversas submuestras mediante reglas univariantes para encontrar la variable independiente que mejor discrimine esta división. Para esto, el algoritmo analizará cada variable explicativa y sus puntos de corte y, finalmente, seleccionará la que mayor homogeneidad aporte a los subgrupos creados. De esta forma se obtienen grupos compuestos por observaciones más homogéneas, incluso más que la que existe en el grupo de procedencia, también llamado nodo madre. Este proceso iterativo finaliza cuando resulte imposible realizar una nueva división que mejore la homogeneidad existente (Puertas Medina, R. y Martí Selva, M., 2013).

Además de los citados, existe un método de aplicación más reciente llamado “redes neuronales”, entendidas como un cúmulo de unidades de cómputo básicas (las neuronas) que conforman múltiples redes paralelas, cuya función y ordenamiento está determinada por la estructura principal. Permiten capturar datos del entorno en forma incremental para convertirlos en información compleja que provean, a su vez, predicciones confiables para nuevas situaciones. Las redes neuronales realizan una diversa gama de tareas, incluyendo la predicción o aproximación de funciones poblacionales, clasificación de patrones, agrupación en clústeres y previsiones. Son asimismo muy potentes para emparejar los datos con el modelo que mejor los explique (Samarasinghe, 2006, p.12).

Llevado al ámbito de los modelos crediticios, las redes neuronales son algoritmos de inteligencia artificial que permiten que el modelo “aprenda” a través de la experiencia para discernir la relación entre las características del prestatario y la probabilidad de incumplimiento. También para determinar qué características son las más importantes para predecirlo. Este método es más flexible que las técnicas estadísticas estándar, ya que no se deben hacer suposiciones sobre la forma funcional de la relación entre las características y la probabilidad de impago o acerca de las distribuciones de las variables o errores del modelo, ya que cuenta con la ventaja de explicar sus correlaciones (Mester, 1997).

2.4. Sistema de información

Tanto los modelos paramétricos como los no paramétricos tienen como insumo fundamental datos acerca del comportamiento del cliente, que se transforma luego en información para la toma de decisiones. Este *output*, por su tenor, es de carácter sensible y está sujeta a reglamentaciones en todo el mundo. Anteriormente se expuso al Comité de Basilea como el marco referencial para el tratamiento de asuntos en esta materia que, en Argentina, está regida por las directivas del Banco Central de la República Argentina.

Según lo normado en el artículo 1° de la Ley 24.144, “el Banco Central de la República Argentina es una entidad autárquica del Estado nacional” (1992), cuya finalidad es “promover, en la medida de sus facultades y en el marco de las políticas establecidas por el Gobierno Nacional, la estabilidad monetaria, la estabilidad financiera, el empleo y el desarrollo económico con equidad social” (Ley 24.144, 1992, art.3).

2.4.1. Métodos de evaluación admitidos

En Argentina, el Banco Central plantea, dentro de los requisitos para el otorgamiento de financiaciones, dos métodos posibles de evaluación para decidir sobre otorgamiento crediticio. El sistema de “*screening*” y los modelos de “*scoring*” (BCRA, 2019).

Se entiende por sistemas de “*screening*” al conjunto de pasos y reglas de decisión que recogen la experiencia acumulada en el otorgamiento de créditos, el seguimiento de su comportamiento posterior y la política de créditos de la entidad.

Este método deberá aplicarse de forma sistemática y actualizarse de manera periódica, a fin de extraer conclusiones en relación con el otorgamiento de créditos y asignar márgenes de financiación.

Por otra parte, los modelos de “*credit scoring*” son métodos matemáticos o estadísticos- econométricos empleados para medir el riesgo y/o la probabilidad de incumplimiento de los solicitantes de crédito.

Ambas técnicas, deben basarse en las variables que las entidades financieras consideren relevantes para medir el riesgo de incobrabilidad asociado a cada deudor y clase de crédito, pudiendo emplear el mismo tipo de información (BCRA, 2019).

2.4.2. Clasificación de deudores

Las entidades financieras, tuteladas por atribución del BCRA, deben clasificar a sus clientes por las financiaciones de acuerdo con el cumplimiento y/o posibilidad de afrontar sus compromisos, en base a una evaluación de su situación particular. Estas incluyen, préstamos, créditos de intermediación financiera y otros créditos, tal lo normado en la sección 2 del texto ordenado para clasificación de deudores (BCRA, 2019).

La sección 5 del citado documento indica que la cartera de clientes se agrupará en comercial y de consumo o vivienda. Esta categoría comprende fundamentalmente el crédito para consumo personal, familiar y tarjetas de crédito. (BCRA, 2019).

Para clasificar a los clientes, el criterio que prima es el de considerar la capacidad de repago del deudor y luego el de liquidación de activos. Para el caso de la cartera de consumo, la calificación es automática en relación con la cantidad de días en mora. Esto se ve reflejado en la sección 6 del mencionado texto, donde también se detallan los seis niveles de clasificación o categorías en la que cada cliente bancario deben encuadrarse.

1: En situación normal. Requiere que cumpla regularmente con el pago de sus obligaciones, con atrasos de hasta 31 días.

2.1: Con seguimiento especial. Requiere que cumpla regularmente con el pago de sus obligaciones, con atrasos de hasta 90 días. Se los considera “de riesgo bajo”.

2.2: En negociación o con acuerdos de refinanciación. Incluye a aquellos que hayan ingresado en refinanciación voluntaria antes de los 60 días de que se verifique la mora.

3: Con problemas. Para clientes con cierta iliquidez, que incurran en atrasos de hasta 180 días o cuenten con refinanciaciones reiteradas. Forman parte de la cartera “de riesgo medio”

4: Con alto riesgo de insolvencia. Para aquellos que incurran en atrasos de hasta un año. También si ha sido demandado judicialmente por alguna entidad para el cobro de su acreencia. Son tratados como clientes “de riesgo alto”.

5: Irrecuperable. Son deudas incobrables, para clientes con mora mayor a un año.

6: Irrecuperable por disposición técnica. Clientes con más de 180 días de atraso que a su vez formen parte de alguna entidad liquidada o en proceso de disolución (BCRA, 2019).

2.4.3. Centrales de información y protección de datos.

De acuerdo con el texto ordenado para centrales de información, los datos mencionados en el punto anterior son recabados por el BCRA y se encuentran a disposición de entidades financieras, proveedores no financieros de crédito y SGR, a través de la publicación mensual de una base de datos actualizada que se coloca en un sitio seguro. La base de datos deberá estar normalizada según se muestra en el anexo II⁸ (BCRA, 2019).

Estos datos son utilizados por bureaus de crédito, centrales de información crediticia que acumulan *insights* sobre el comportamiento de cada individuo y sirven a su vez como fuente de consulta para las entidades del sistema financiero.

La información suministrada por las centrales de información está respaldada por la ley 25.326 de protección de datos personales, la que en su artículo 26 regula el tipo de dato que puede alojarse en sus bases:

1. En la prestación de servicios de información crediticia sólo pueden tratarse datos personales de carácter patrimonial relativos a la solvencia económica y al crédito, obtenidos de fuentes accesibles al público o procedentes de informaciones facilitadas por el interesado o con su consentimiento.
2. Pueden tratarse igualmente datos personales relativos al cumplimiento o incumplimiento de obligaciones de contenido patrimonial, facilitados por el acreedor o por quien actúe por su cuenta o interés.
3. A solicitud del titular de los datos, el responsable o usuario del banco de datos, le comunicará las informaciones, evaluaciones y apreciaciones que sobre el mismo hayan sido comunicadas durante los últimos seis meses y el nombre y domicilio del cesionario en el supuesto de tratarse de datos obtenidos por cesión.
4. Sólo se podrán archivar, registrar o ceder los datos personales que sean significativos para evaluar la solvencia económico-financiera de los afectados durante los últimos cinco años. Dicho plazo se reducirá a dos años cuando el deudor cancele o de otro modo extinga la obligación, debiéndose hacer constar dicho hecho.

⁸ La base es de acceso público en: [https:// www.bcra.gob.ar/BCRAyVos/Situacion_Crediticia.asp](https://www.bcra.gob.ar/BCRAyVos/Situacion_Crediticia.asp)

5. La prestación de servicios de información crediticia no requerirá el previo consentimiento del titular de los datos a los efectos de su cesión, ni la ulterior comunicación de ésta, cuando estén relacionados con el giro de las actividades comerciales o crediticias de los cesionarios (Ley 25.326, 2000).

2.5. Inclusión financiera

Con lo expuesto hasta este punto se puede vislumbrar un circuito en el que la experiencia crediticia de cada cliente alimenta sistemas de información, que permiten predecir su comportamiento futuro, lo que habilita a este individuo a contar con una nueva posibilidad de seguir tomando dinero del sistema financiero, en un círculo virtuoso. Esto aplica para aquellos que se encuentran dentro del sistema pero ¿qué pasa con los que no forman parte? ¿Cómo pueden incluirse? ¿De qué hablamos cuando hablamos de inclusión financiera?

2.5.1. Definición y limitantes

De acuerdo con el Banco Mundial (2018), se entiende por inclusión financiera a permitir el acceso a productos financieros útiles y asequibles, prestados de manera responsable y sostenible. No solo a una cuenta bancaria, la que se ve como una puerta de acceso a servicios financieros más sofisticados.

Además, el organismo postula que estas herramientas permiten iniciar y ampliar negocios, generar inversiones y paliar inconvenientes financieros, lo que redundaría en una mejor calidad de vida para la persona y económica para el conjunto de la sociedad en la que participa (Banco Mundial, 2018).

Entre las barreras a la inclusión financiera se pueden citar, siguiendo los lineamientos del G20, ingresos insuficientes de la población, poca educación financiera y dificultad de acceso a las sucursales (G20, 2018, p.2). Fundamentalmente la informalidad contrae de forma significativa la inclusión financiera. Los agentes económicos que no registrados tienen severas dificultades para acceder a los servicios financieros formales. En particular por el criterio de selección de la oferta, que necesita documentación con la que los informales no poseen por definición (G20, 2018, p.3).

Dentro de los aspectos legales que limitan la inclusión, se puede mencionar el asegurar la validez legal de los documentos de identidad a bajo costo y el establecer reglas de juego claras que protejan a los usuarios (Banco Mundial, 2018).

Por lo expuesto, la digitalización de la economía puede permitir agrandar la base de incluidos en el sistema financiero, con herramientas que derriben las barreras de elegibilidad, físicas y de costo del crédito (G20, 2018, p.3).

2.6. Entorno Fintech: Más que una moda

¿Cómo se podría abordar una solución a los problemas citados en los apartados anteriores? La aplicación de estas herramientas en un entorno *Fintech* podría ser una respuesta a este interrogante.

2.6.1. Definición y pertinencia

Su nombre se basa en un neologismo que proviene de la unión de Finanzas y Tecnología. En un sentido más amplio, siguiendo la definición vertida por la consultora Ernst & Young (2017)

Definimos a las *Fintech* como organizaciones de alto crecimiento que combinan modelos de negocio innovadores y tecnología para permitir, mejorar e irrumpir en la industria financiera. Esta definición no se limita a las *start-ups* ni a los nuevos operadores, sino que incluye a *scale-ups*, empresas en etapa de maduración e incluso empresas ajenas a los servicios financieros, como los proveedores de telecomunicaciones y los minoristas electrónicos (p.25).

La irrupción del *Fintech* y el aumento del uso de teléfonos móviles han facilitado el acceso a servicios financieros a poblaciones difíciles de atender, con un costo más bajo y con menos riesgos. Por ejemplo, los documentos digitales facilitaron la apertura de cuenta. El dinero virtual mejoró la transacción entre individuos y los datos recolectados por estos nuevos sistemas permite a los proveedores financieros diseñar productos a medida de los no-bancarizados (Banco Mundial, 2019).

Argentina es un país fértil para el desarrollo tecnológico. Cuatro de las más grandes⁹ empresas de base tecnológica han surgido en este país. Por ello, no es de extrañar que algunas herramientas utilizadas en el entorno *Fintech* ya estén dando muestras de utilidad en la región.

Se ha popularizado el uso de *Big Data*, con algoritmos avanzados que permiten que se procese y sintetice la información recolectada. A punto tal que Mercado Libre, por

⁹ Llamadas unicornios, son empresas con valuación bursátil mayor a mil millones de dólares. Para el periodo analizado las cuatro organizaciones de origen local eran Mercado Libre, Despegar.com, OLX y Globant.

intermedio de su compañía Mercado Crédito, ha salido a ofrecer financiamiento a sus usuarios, a los que califica por su historial de comportamiento en la empresa, gracias a las técnicas de minería de datos (Mercado Crédito, 2017). El Banco de la Nación incentiva el envío de dinero y pagos a través del celular con su plataforma PIM. La reciente presentación de TransferWise presupone un avance en servicios P2P¹⁰.

2.6.2. *Redes sociales*

El mayor maridaje entre información personal y tecnología queda a cargo de las redes sociales, entendidas según Boyd y Ellison (2007, p.211) como un servicio basado en la web que permite a los individuos articular una lista de conexiones con otros usuarios dentro del sistema y la naturaleza de estas conexiones puede variar entre cada prestador.

Una vez que una persona se une a una red social, completa una serie de datos personales. Generalmente se pregunta la edad, localización e intereses personales. Adicionalmente se suele solicitar una foto de perfil. El nivel de privacidad depende tanto del nivel de discrecionalidad que decida cada usuario y del nivel de privacidad que ofrezca la plataforma (Boyd y Ellison, 2007, p.213).

En el caso de Twitter, se define como una red social de carácter público. Los tuits – mensajes publicados – e información de perfil están a disposición de cualquier motor de búsqueda (Twitter, 2019). En su política de privacidad, la herramienta estipula que

la mayor parte de la actividad en Twitter es pública, lo que incluye su información de perfil, su zona horaria e idioma, la fecha de creación de su cuenta y sus tuits, así como cierta información de sus tuits como la fecha, hora y la aplicación y versión de Twitter desde la que tuiteó. También puede decidir publicar su ubicación en sus tuits o en su perfil de Twitter. Las listas que crea, la gente a la que sigue y que le sigue, así como los tuits a los que hace me gusta o retuitea también son públicos. Las transmisiones de Periscope que usted crea, en las que hace clic o en las que participa de otra forma, ya sea en Periscope o Twitter, son públicas junto con la información sobre cuándo realizo dichas acciones (Twitter, 2019).

¹⁰ P2P o *peer to peer* refiere a tecnología que une punta a punta oferentes con demandantes, sin intermediarios.

2.6.3. *Mailing*

El envío de correo electrónico configura un subtipo de comunicación diferente al de las redes sociales. Definido como “buzoneo” por la RAE, es una evolución del “envío masivo de correspondencia personalizada realizado por una empresa o una institución” (RAE, 2019).

Dentro de sus ventajas se puede destacar que es fácil de implementar y de medir, además de ser automatizable y eficiente. Como desventaja se destaca la sobrecarga de mensajes recibidos, lo que dificulta sobresalir sobre los demás.

Para llevar adelante esta acción, es necesario contar con una base de suscriptores que se registran voluntariamente pero, del mismo modo, pueden de suscribirse cuando lo deseen. Forma parte de las buenas prácticas en esta disciplina el revisar la lista de inscriptos periódicamente en busca de direcciones repetidas, erróneas o fallidas (Fariborzi, E. y Zahedifard, M, 2012).

Capítulo III Modelización tradicional

De acuerdo con lo planteado en el marco teórico se realizará, a lo largo de este capítulo, una modelización tradicional de scoring, con el objetivo de exponer el poder predictivo de las herramientas que generalmente utiliza el sistema financiero para predecir el comportamiento crediticio de sus clientes. De esta manera, se define en este capítulo un modelo que funcionará como *baseline* de comparación al ser contrastado contra la nueva metodología propuesta en el capítulo IV, que incorpora variables del entorno *FinTech*.

3.1. Universo analizado

La modelización tradicional se realizó sobre una base de datos del Banco Columbia, entidad de crédito de capitales nacionales con atención predominantemente minorista, para el período enero-2016 a junio-2018. Dicha ventana temporal responde a que se debe verificar, en todos los casos, que cada préstamo presente un recorrido de 12 meses desde el otorgamiento para el análisis. La base de datos general contaba con 355.744 observaciones sobre las que se aplicaron ciertas restricciones acordes al objetivo de esta investigación. Se filtraron solo los créditos otorgados cuyos clientes hayan completado en la solicitud el correo electrónico, independientemente de la veracidad del dato. Esta condición se debe a que el e-mail constituye uno de los diferenciales que serán utilizados al momento de desarrollar un modelo basado en entorno *FinTech*. La base de datos resultantes fue de 4.580 préstamos que resultan en el set de datos a razón de uno por persona.

Se seleccionaron distintas variables explicativas del comportamiento de los clientes en el mercado de crédito siguiendo lo expuesto en el marco teórico previamente detallado y se construyeron algunas otras. A los fines de explicar el sentido de su elección, fueron agrupadas en sociodemográficas, de producto, y de comportamiento crediticio. El detalle de su tipología puede encontrarse en la tabla I del anexo III, mientras que los principales indicadores de estadística descriptiva se desarrollan en la tabla II del mismo apartado.

3.1.1. Variables sociodemográficas

Bajo esta categoría se engloban las variables necesarias para entender quién es el cliente y cuál es su nivel de ingreso. Se tiene en cuenta la variable “SEXO”, en donde se marca el género del solicitante del préstamo. Esta variable toma especial relevancia en un contexto de histórica disparidad de ingresos entre hombres y mujeres. Se incluye también la

variable “Edad” al momento de tomar el préstamo que se relaciona también con la expectativa de vida. Para el set de datos se verifica un 61% de mujeres y un 39% de hombres, con una edad promedio de 57 años y una moda de 60 años. A priori, se puede inferir que son personas que están en el declive de su ciclo de vida laboral, toda vez que 60 años era la edad jubilatoria para las mujeres al momento de realizarse este análisis. El rango etario 21 a 84 años se enmarca en la política de préstamos del Banco, dado que, a partir de los 85 años, los tomadores de créditos no son pasibles de ser asegurados con seguros de vida por saldo deudor.

Con la “Etapa laboral”, se hace foco en el ciclo de vida del prospecto al abordar la característica de “activos” (variable = 1) y “pasivos” (variable = 0), para entender si está en una etapa incipiente o decreciente en su capacidad de generación de ingresos. Esta información se complementa por medio de la variable “Actividad”, que categoriza a cada persona por actividad económica, por ejemplo: empleado privado, empleado público, jubilado, pensionado, etc. A su vez, éstas están distribuida a su vez en una serie de variables *dummies* que indican si el cliente fue marcado con una u otra.

La combinatoria de este subgrupo de variables, aplicadas al objeto de estudio, muestra que el 53% de los individuos están catalogados como pasivos, ya sean jubilados (3) o cuenten con alguna pensión (7) y un 24% adicional se enrola como empleado privado. Si se suma el 7,5% que aduce ser empleado municipal, se llega a la conclusión que el 85% de los clientes tienen un ingreso mensual habitual.

Luego se controla por variables de ingreso, siendo “IngresoUsado” un indicador del ingreso inferido o real para cada cliente, que luego se corrobora con una variable *dummy* de “Presenta recibo”, en función si justifica ese ingreso. El ingreso promedio de los tomadores de préstamos, en términos nominales, fue de \$14.000 en el período de estudio. Como referencia el SMVM promedio del período fue de \$8.155

Finalmente se controla por procedencia geográfica del tomador de crédito mediante las variables “ProvinciaID”, “Region” y “Sucursal”. La muestra indica que uno de cada dos clientes proviene de la zona denominada AMBA, área metropolitana de Buenos Aires, dato que condice con la distribución censal de la población.

3.1.2. Variables de producto

Este conjunto de variables se relaciona con el tipo de producto crediticio préstamo de consumo, tendiente a satisfacer necesidades de financiamiento puntuales de personas físicas, o jurídicas que se encuentren expuestas en menos de \$500.000 en el sistema financiero.

La variable del modelo **“Producto sistema financiero”** es una construcción propia del Banco que determina el tipo de préstamo, el método de originación y la forma de cobranza. Así por caso se pueden encontrar, siempre codificados, préstamos para jubilados contactados por teléfono cuyo repago sea a través del débito automático (77) como asistencia a activos contactados en sucursal con repago vía pago voluntario con cuponera (13). Precisamente este producto es el más requerido, con un 48% de solicitudes, según se desprende del set de variables *dummies* generadas para esta característica.

La variable **“Tipo de pago”** separan entre quienes adoptan por un repago en forma voluntaria de quienes comprometen un débito automático mensual. Sólo un 26% de la muestra permite el débito, lo que permite inferir que el comportamiento de cada prospecto y su voluntad de pago influirá en la probabilidad de cancelar el crédito en tiempo y forma. La variable **“Tipo de operación”**, por su parte, indica si se trata de un préstamo primario o una renovación. Dado que casi el 80% de la muestra aplica para préstamos primarios, se tiene menor conocimiento del comportamiento crediticio del cliente.

El **“Capital solicitado”** es el monto puro de la operación, sin contar impuestos ni gastos. La duración del préstamo, por su parte, se visualiza en el **“Plazo (meses)”**. Durante el período analizado, el capital solicitado promedio nominal fue de \$39.000, pero la mayoría de los préstamos se concentró en el orden de los \$10.000, con 701 clientes.

Analizando la duración de la relación crediticia, se verifica que los plazos están dentro de lo admitido por las políticas de crédito del Banco, que inicia en 6 meses y tiene un tope de 5 años. El promedio observado es de 28 meses, que a su vez se ajusta a la moda de 24 meses de compromiso.

La relación entre el capital y los ingresos del cliente definen el **“Ratio capital ingreso”**, variable construida a los fines de enriquecer el modelo y mitigar la nominalidad del dinero expresado en distintos momentos del tiempo. Para la muestra actual, se verifica un ratio promedio de 3,8 veces capital pedido por sobre el ingreso del cliente.

3.1.3. Variables de comportamiento crediticio

Por último, se consideran una serie de variables en relación a la historia crediticia del cliente. La fuente de la información es el BCRA y los bureaus de crédito contratados a tal efecto. En esta categoría se encuentran variables de exposición, expresadas mediante saldos de deuda nominal en un período de tiempo, de cantidad de entidades con las que opera el posible tomador de crédito y de situación BCRA para distintos momentos del tiempo, de forma tal que se pueda construir un historial crediticio.

Cabe aclarar en este punto que en el set de datos provisto por el Banco Columbia no se encuentran clientes con situación BCRA actual mayor a 2 porque incorporar a la cartera a un individuo que ya posea una deuda con atraso mayor a 90 días (situación 3 o más, vale decir, defaulteada) implicaría generar una previsión contable que atenta contra la rentabilidad del negocio. No obstante, esta restricción rige bajo normativa del BCRA al momento del análisis y es análoga a todas las entidades, lo que mitiga la incidencia de esta variable. Eso limita levemente las conclusiones generales del modelo.

3.1.4. Variable objetivo

A la base originada se agregó una columna denominada `flag_PD12`, variable de tipo dicotómica que asigna el valor 1 a todo cliente que haya pasado más de 90 días sin realizar un pago de cuota total durante los primeros 12 meses de la vida del préstamo. Esta es la variable que se querrá predecir con el modelo.

3.2. Análisis bivariado:

A continuación se procede a realizar un análisis bivariado de cada una de estas variables presentadas anteriormente sobre la variable objetivo. Por ejemplo: Que el cliente tenga sexo femenino o masculino ¿afecta su probabilidad de incurrir en default?, o dicho en forma más taxativa: ¿Cuál es el promedio de clientes sexo femenino y masculino que hayan caído en default en los últimos doce meses (`flag_PD12 = 1`)? ¿Existe algún patrón de comportamiento que asocie esa característica con la variable objetivo? Para responder a estas preguntas, es necesario realizar un análisis bivariado de cada una de las variables consideradas, cada una en forma independiente, como para tener un primer filtro estadístico que permita determinar si es relevante incorporarlas o no al modelo que se pretende construir. Asimismo, esta herramienta facilita un agrupamiento de las variables ordinales y categóricas que permite identificar patrones de comportamiento por rangos.

Nuevamente, para mantener la cohesión del texto, se agrupan en variables sociodemográficas, de producto, y de comportamiento crediticio y se acompaña el detalle pormenorizado en las tablas III, IV y V del anexo III.

3.2.1. Variables sociodemográficas

Este set de características brinda algunas definiciones consistentes. En primera medida, la variable “**SEXO**” indica que las mujeres de la muestra tienen mejor comportamiento crediticio que los hombres, con una tasa de default 15% más baja. Haciendo foco en el rango etario, la probabilidad de que un cliente no pague sus créditos es descendente a medida que envejece, ya sea medido por “*Edad*” como por la “*Etapas laborales*” en la que se encuentra. En efecto, se produce un marcado corte en el comportamiento de la variable cuando se llega al umbral de los 65 años, edad jubilatoria máxima para mujeres y mínima para hombres.

Si se mide el impacto de realizar determinada “*Actividad*”, se verifica del mismo modo que aquellas tareas asociadas a la vida pasiva tienen mayores ratios de default que las enmarcadas en el ciclo de vida activo.

Contra intuitivamente, el “*Ingreso Usado*”, indicador del ingreso asumido para cada cliente, reacciona en forma directa. Vale decir que, a menor ingreso considerado para el individuo, menor probabilidad de ingresar en cesación de pagos. Mismo comportamiento para cada cliente que “*Presenta recibo*”, lo que permite reforzar la idea que cuanto más certero es el valor a considerar, más se verifica esta linealidad entre ingresos y tasa de default.

A nivel geográfico, analizado por “*Region*”, la zona de AMBA muestra un peor desempeño que el observado en clientes del interior del país. Como hipótesis que valide estos guarismos puede aducirse que las sucursales del interior trabajan con concepto de cercanía y conocimiento del cliente mucho mayor que el que se da en el área metropolitana.

3.2.2. Variables de producto

Dado que el producto es una construcción propia del Banco, el “*Producto sistema financiero*” (PSF) refleja en buena medida la eficiencia de las políticas de originación de cada uno de ellos, siempre considerando que cada producto está condicionado con el tipo de actividad económica del cliente. Por tal motivo no es de extrañar que los PSF asociados a clientes activos tengan peor comportamiento que los relacionados a pasivos. Sin embargo, los ratios de default son más altos en algunos casos, siendo el PSF 13 el más emblemático, con un índice de 3x entre buenos y malos pagadores. Esto se debe a que también debe

considerarse la forma en la que se cancela el préstamo, sea voluntaria o por débito. Los individuos del PSF 13 son clientes de pago voluntario por cuponera y, por lo que se desprende del análisis bivariado sobre la variable **“Tipo de pago”**, solo el 0,17% de los clientes que tienen débito automático ingresaron en default, contra el 5,73% de los casos de pago voluntario. Números contundentes que tienen una explicación racional: es muchísimo más probable cobrarle a un cliente que tiene asociada una cuenta con débito automático que a uno que debe presentarse voluntariamente a pagar todos los meses.

Volviendo sobre las características del producto, el **“Capital solicitado”** demuestra cierta lógica a la hora de medir comportamiento. Cuanto más se pide, más chance de que se entre en default, pero al compararlo con el ingreso mediante el **“Ratio capital ingreso”** se puede observar que, a pesar de que los clientes tienen más dificultad para pagar a medida que el capital comprometido equivale a mayor cantidad de ingresos, pueden hacer frente a ratios muy altos de capital/ingresos. Esto se explica basado en que algunos clientes con bajos ingresos, producto de asignaciones asistenciales, toman préstamos de bajo monto absoluto pero alto relativo, mediante la modalidad de pago con débito, lo que reduce fuertemente su probabilidad de impago.

Por último, se verifica con el análisis de la variable **“Plazo (meses)”** que los préstamos más cortos (de hasta 18 meses) tienen mejor performance que los compromisos más largos, en el orden del 15% de mejora. El **“Tipo de operación”**, por otra parte, no permite diferenciar buenos de malos pagadores.

3.2.3. Variables de comportamiento crediticio

Este set de variables permite entender la relación entre el comportamiento en el sistema financiero y la probabilidad de que vaya a pagar un crédito al banco analizado. Haciendo foco en la situación crediticia al momento de la calificación, se observa que los clientes que tenían **“Exposicion”** previa dentro del sistema financiero son casi 30% peores pagadores que quienes no tenían compromisos, que la tasa de defaulteados sube acorde a la **“Exposicion actual BCRA”** medida en pesos nominales y que el **“Ratio exposicion ingreso”** mejora las probabilidades de repago si el índice es menor al 100%.

Al realizar una separación los clientes según la cantidad de bancos con los que opera, con la variable **“Cantidad Bancos actual”**, se puede concluir que la probabilidad de default aumenta para aquellos clientes bancarizados, hasta llegar a un límite de más de 5 bancos, momento en el que baja significativamente, debido a que la sexta entidad y subsiguientes refuerzan sus políticas de crédito y solo le prestan dinero a quien realmente confían que se

lo pueda devolver, por lo que el agrupamiento categórico se dividió entre quienes no operaban con ningún banco, los que lo hacían solo con uno y los que tenían compromisos con más de una entidad. El resto de las características de comportamiento actual están fuertemente sesgadas por las políticas de otorgamiento de Banco, motivo por el cual no se encuentran clientes con varias referencias comerciales o calificados con situación BCRA mayor a 2, lo que les quita materialidad.

Avanzando en el comportamiento histórico y medido el estatus crediticio a seis meses, encontramos sentido económico en la relación entre las variables independientes y la objetivo, la que demuestra subas constantes en la probabilidad de ocurrencia a medida que se deterioran los indicadores. Dado que el horizonte semestral se considera “de situación reciente”, las variables de calificación crediticia que lo componen también están condicionadas por las políticas de originación.

A doce meses, la situación se mantiene sin modificaciones. Las variables de situación BCRA no tienen aún un recorrido tan largo como para cambiar el impacto en el comportamiento del cliente, aunque comienza a reflejarse una mínima tendencia, no significativa, en la relación entre cantidad de bancos en situación normal a doce meses (“*Cant sit 1 12m*”) y la tasa de default.

Por último, al analizar el comportamiento del cliente a veinticuatro meses se verifican las tendencias marcadas a seis y doce meses, con ligeras diferencias. Por caso, la curva de cantidad de bancos operados en el periodo (“*Bancos 24m*”) tuerce su forma a partir de la séptima entidad, por lo que también se agrupó entre quienes no registran comportamiento crediticio a dos años y los que si lo han tenido. Asimismo, adquiere materialidad la variable que reconoce las situaciones BCRA mayores a 2 (“*Cant sit >2 24m*”), con una curva de tendencia lógica que asocia la cantidad de entidades en este estado con la probabilidad de incumplimiento del préstamo.

3.3. Modelo tradicional

El análisis bivariado permitió reacondicionar el set de variables originales, quitar las que no aportaban al objetivo de predecir el comportamiento crediticio y transformar aquellas de tipo continuas a categorías que respeten uniformidad u ordenamiento, por ejemplo “*Edad*”, o bien, para las que no se espera que ordenen, que tras la categorización demuestren comportamiento tipo “serrucho”. Realizada esta tarea, se procedió a renombrar las variables

trabajadas con el sufijo “_S”, de forma de indicar que ya no se trata de un dato en crudo, sino de uno tamizado.

Se utilizó el software estadístico SPSS Modeler para el procesamiento de datos. Se incorporó toda la información previamente refinada a una ruta de trabajo, cuya metodología de tratamiento por nodos, o pasos, resulta conveniente para explicar cómo se obtuvo el modelo predictivo tradicional.

Como primera etapa, se pasó a la partición de los datos en dos grupos, uno de entrenamiento y otro de validación, seleccionando para esto una muestra aleatoria compuesta por el 80% de la base para entrenamiento y un 20% restante para validación.

El modelo predictivo se construyó a partir de una regresión logística binaria, en el que se pretende predecir la variable independiente “*Flag_pd_12*”, que señala a aquellos clientes que incurren en default en un año, a través de una serie de variables explicativas del fenómeno.

Se postula un ajuste y estimación de parámetros contrastados a través de la razón de verosimilitud, con un umbral de significación de 0,05 para la entrada de nuevas variables, debido a que por estándar de industria se toma un intervalo de confianza al 95% y 0,1 para la salida, dado que ésta tiene que ser mayor que el atributo de entrada. Las variables se incorporarán en simultáneo, para luego medir su poder predictivo y significación de variables e ir retirando la que menos ajuste al modelo, hasta llegar a la ecuación que mejor pueda predecir la variable objetivo. Por último, dado que la tasa de default promedio para los clientes del banco se encontraba en el orden del 5,9%, se consideró ese valor como tasa de corte.

3.3.1. Variables del modelo

De las condiciones antes explicadas se desprende un modelo cuyas variables incorporadas fueron “*Edad_S*”, “*Actividad 1_S*”, “*Region_S*”, “*Tipo de pago_S*”, “*PSF13_S*”, “*Ratio capital ingreso_S*”, “*Plazo (meses)_S*”, “*Cantidad Bancos actual_S*”, “*Bancos 12m_S*”, “*Sit BCRA 24m_S*” y “*Cant sit >2 24m_S*”. Entonces, la probabilidad de default aplicable a cada cliente responde a la siguiente ecuación:

$$PD = \frac{1}{1 + e^{(2,136 + 0,352 * Edad_S(1) + 0,56 * Edad_S(2) + 0,421 * Actividad_1_S(1) + 0,309 * Region_S(1) + 3,61 * Tipo_de_pago_S(1) - 0,564 * PSF13_S(1) - 0,035 * Ratio_capital_ingreso_S(1) - 0,566 * Plazo_(_meses)_S(1) + 0,724 * Cantidad_Bancos_actual_S(1) + 0,087 * Cantidad_Bancos_actual_S(2) + 1,258 * Bancos_12m_S(1) - 0,474 * Sit_BCRA_24m_S(1) - 0,215 * Cant_sit > 2_24m_S(1))}}$$

De esta fórmula se desprende el Score, entendido como $(1 - PD) * 1000$.

Su impacto medido en términos de β , los grados de libertad asociados a la cantidad de instancias de cada una (expresados como n-1) y su prueba de significancia medida a través del test de Wald se detallan en la tabla VI.

Tabla VI - Resultado del modelo

| VARIABLES | B | Error estándar | Wald | gl | Sig. |
|------------------------------------|--------|----------------|--------|----|-------|
| <i>Edad_S</i> | | | 4,928 | 2 | 0,085 |
| <i>Edad_S(1)</i> | -0,352 | 0,187 | 3,551 | 1 | 0,06 |
| <i>Edad_S(2)</i> | -0,56 | 0,3 | 3,492 | 1 | 0,062 |
| <i>Actividad 1_S(1)</i> | -0,421 | 0,196 | 4,585 | 1 | 0,032 |
| <i>Region_S(1)</i> | -0,309 | 0,168 | 3,395 | 1 | 0,065 |
| <i>Tipo de pago_S(1)</i> | -3,61 | 0,747 | 23,37 | 1 | 0 |
| <i>PSF13_S(1)</i> | 0,564 | 0,239 | 5,567 | 1 | 0,018 |
| <i>Ratio capital ingreso_S(1)</i> | 0,035 | 0,21 | 0,028 | 1 | 0,867 |
| <i>Plazo (meses)_S(1)</i> | 0,566 | 0,199 | 8,139 | 1 | 0,004 |
| <i>Cantidad Bancos actual_S</i> | | | 6,371 | 2 | 0,041 |
| <i>Cantidad Bancos actual_S(1)</i> | -0,724 | 0,337 | 4,62 | 1 | 0,032 |
| <i>Cantidad Bancos actual_S(2)</i> | 0,087 | 0,68 | 0,016 | 1 | 0,898 |
| <i>Bancos 12m_S(1)</i> | -1,258 | 0,624 | 4,069 | 1 | 0,044 |
| <i>Sit BCRA 24m_S(1)</i> | 0,474 | 0,258 | 3,379 | 1 | 0,066 |
| <i>Cant sit > 2 24m_S(1)</i> | 0,215 | 0,304 | 0,501 | 1 | 0,479 |
| <i>Constante</i> | -2,136 | 0,384 | 30,958 | 1 | 0 |

Al observar los resultados de β , se puede inferir cual será la relación entre poseer determinada característica y la posibilidad de *default* dado que las variables con signo negativo muestran rasgos tendientes a clientes con menor probabilidad de impago y las que tienen signo positivo más proclives a aumentar el ratio de morosidad, todo esto sujeto a la codificación de las variables provista por el software estadístico, que se detalla en tabla VII.

Por caso, a medida que se avanza en edad, el signo negativo se acentúa, en forma análoga a lo explicado en el análisis bivariado. Los adultos mayores son mejores pagadores en este modelo.

Por otra parte, se observa nítidamente que tener la atribución “Tipo de pago” reduce la posibilidad de impago. Esto implica que la variable sea igual a 1 y corresponda a clientes que paguen por débito. Dicho más claramente: los clientes que estén suscriptos al débito automático de sus cuotas serán mejores pagadores que quienes paguen en forma voluntaria. También sucede con los clientes con actividad 1 (empleado privado) en desmedro de las demás actividades, y con los que están radicados en el interior del país por sobre los que realizaron operaciones en AMBA.

Con relación al comportamiento en el sistema crediticio, tener actividad crediticia con un banco hace más fiables a los clientes que quienes no tienen ninguna, pero menos que quienes computan más de una entidad fiadora. Avanzando en el historial, registrar antecedentes en los últimos doce meses mitiga la probabilidad de default, pero haber caído en atrasos de más de 30 días (situación 2 BCRA) en los últimos dos años empeora la situación del cliente.

Tabla VII - Codificaciones de variables categóricas

| Variable | R | Frec | Codific param | |
|--------------------------------|-----|------|---------------|----|
| | | | -1 | -2 |
| <i>Edad_S</i> | 1.0 | 959 | 0 | 0 |
| | 2.0 | 1683 | 1 | 0 |
| | 3.0 | 1012 | 0 | 1 |
| <i>Actividad 1_S</i> | 0.0 | 2769 | 0 | |
| | 1.0 | 885 | 1 | |
| <i>Tipo de pago_S</i> | 0.0 | 2721 | 0 | |
| | 1.0 | 933 | 1 | |
| <i>Cant sit >2 24m_S</i> | 0.0 | 3143 | 0 | |
| | 1.0 | 511 | 1 | |
| <i>Bancos 12m_S</i> | 0.0 | 972 | 0 | |
| | 1.0 | 2682 | 1 | |
| <i>Ratio capital ingreso_S</i> | 0.0 | 2481 | 0 | |
| | 1.0 | 1173 | 1 | |

| Variable | R | Frec | Codific param | |
|---------------------------------|-----|------|---------------|----|
| | | | -1 | -2 |
| <i>Cantidad Bancos actual_S</i> | 1.0 | 484 | 0 | 0 |
| | 2.0 | 680 | 1 | 0 |
| | 3.0 | 2490 | 0 | 1 |
| <i>Region_S</i> | 0.0 | 1863 | 0 | |
| | 1.0 | 1791 | 1 | |
| <i>PSF13_S</i> | 0.0 | 1852 | 0 | |
| | 1.0 | 1802 | 1 | |
| <i>Sit BCRA 24m_S</i> | 0.0 | 2804 | 0 | |
| | 1.0 | 850 | 1 | |
| <i>Plazo (meses)_S</i> | 0.0 | 1122 | 0 | |
| | 1.0 | 2532 | 1 | |

Tal como se explicara previamente, algunas variables se mantuvieron en el modelo a pesar de no ser estadísticamente significativas ($\text{sig} > 10\%$). El “*PSF13_S(1)*”, indicativo de clientes que se encuentren en etapa activa y paguen por cuponera, por su importancia a la hora de caracterizar al perfil más riesgoso, en contraposición al pasivo que paga por débito. En similar sentido el “*Ratio capital ingreso_S(1)*”, que encuentra sentido al identificar clientes que hayan pedido préstamos que excedan en más de cinco veces sus ingresos y

“*Cant sit >2 24m_S(1)*” indispensable para saber si los clientes tuvieron atrasos en sus compromisos en los últimos dos años.

En forma complementaria, el cuadro reflejado en la tabla VIII muestra el poder predictivo de cada variable medido en forma individual a través de sus respectivos KS, GINI y AUC, verificándose que cada una de ellas tiene un área bajo la curva mayor al 50%, por lo que las hace preferente al azar.

Tabla VIII – Poder predictivo individual

| Variable | KS | GINI | AUC |
|--------------------------------|----|-------|-------|
| <i>Edad</i> | 17 | 22,78 | 61,39 |
| <i>Actividad 1</i> | 8 | 8,89 | 54,45 |
| <i>Tipo de pago</i> | 25 | 25,43 | 62,71 |
| <i>Cant sit >2 24m_S</i> | 5 | 5,39 | 52,7 |
| <i>Bancos 12m_S</i> | 0 | 0,06 | 50,03 |
| <i>Ratio capital ingreso_S</i> | 9 | 9,69 | 54,85 |

| Variable | KS | GINI | AUC |
|---------------------------------|----|-------|-------|
| <i>Cantidad Bancos actual_S</i> | 3 | 3,85 | 51,93 |
| <i>Region_S</i> | 5 | 5,43 | 52,71 |
| <i>PSF13_S</i> | 26 | 26,55 | 63,28 |
| <i>Sit BCRA 24m_S</i> | 8 | 8,6 | 54,3 |
| <i>Plazo (meses)_S</i> | 5 | 5 | 52,5 |

3.4. Pruebas de estimador

Una vez determinado el modelo que predice la probabilidad de impago, se debe probar que tenga relevancia estadística, a través de una batería de herramientas.

3.4.1. Matriz de clasificación

El primer indicador para saber si un modelo logra predecir correctamente consiste en contrastar los resultados obtenidos por el modelo contra el resultado real visualizado en cada cliente. Así se tendrá una matriz de clasificación (tablas IX y X), que tomará como valor pronosticado a la probabilidad de que un cliente caiga en default, utilizando la tasa de corte de 0,059 anteriormente declarada, asignando el valor 1 a quienes superen ese rango.

Tabla IX - Tabla de clasificación

| Observado | | | Pronosticado | | |
|-------------------|---------------|-----|---------------|-----|---------------------|
| | | | Flag_PD12_Max | | Porcentaje correcto |
| | | | 0.0 | 1.0 | |
| Paso 1 | Flag_PD12_Max | 0.0 | 2571 | 921 | 73,6 |
| | | 1.0 | 64 | 98 | 60,5 |
| Porcentaje global | | | 3654 | | 73,0 |

Tabla X - Matriz de clasificación (en %)

| Observado / Pronosticado | Positivo | Negativos | Totales |
|-----------------------------|----------|-----------|---------|
| Positivos | 2,68% | 1,76% | 4,44% |
| Negativos | 25,20% | 70,36% | 95,56% |
| Totales | 27,88% | 72,12% | 100,00% |

De las tablas precedentes se desprenden otros indicadores, como lo son Precision y Recall Rate,

$$\text{Precision Rate: } \frac{2,68}{27,88} = 9,61\%$$

$$\text{Recall Rate: } \frac{2,68}{4,44} = 60,36\%$$

De los estos valores se concluye que el 9,61% de los clientes sea marcados como default, y el 60,36% de los casos positivos totales fueron clasificados correctamente por el modelo.

3.4.2. Scorecard

Una vez calculado el Score para cada cliente de la muestra de entrenamiento, se puede realizar una separación por rangos que permita por un lado encontrar un ordenamiento entre scores más bajos a mayor cantidad de defaulteados y, consecuentemente, observar si existe una buena discriminación entre la distribución de buenos y malos pagadores en cada rango.

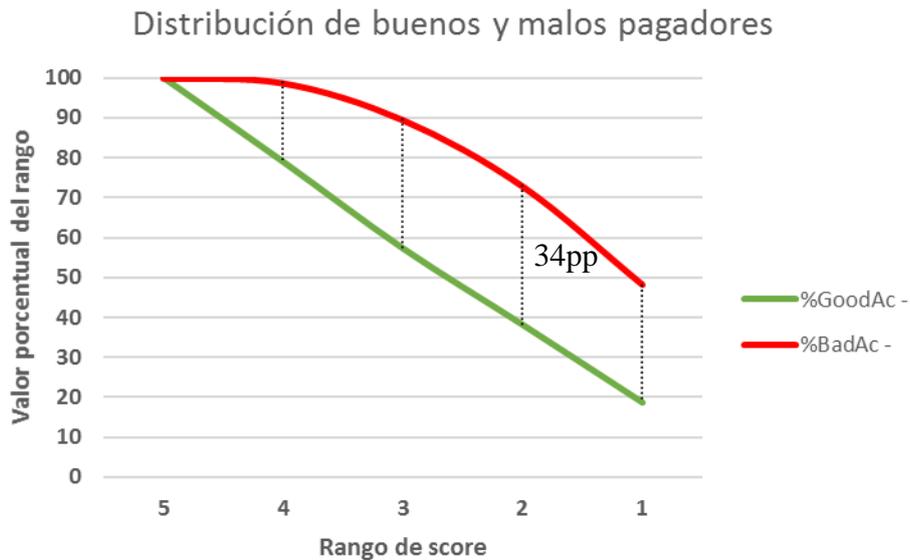
La scorecard que se detalla en tabla XI surge, entonces, de la separación en quintiles de los clientes analizados en entrenamiento. Se establece en 5 rangos debido a la cantidad de casos presentes cuando se haga la validación.

Tabla XI – Scorecard población entrenamiento

| Rango Score | Score Mín. | Score Máx. | #Total | %Good Ac - | %Bad Ac - | %Bads | PD Teórica | KS | HL |
|----------------|---------------|---------------|--------|---------------|--------------|-------|---------------|----|---------------|
| 5 | 997 | 999,82 | 737 | 100 | 100 | 0,27 | 0,16 | 19 | 0,3312 |
| 4 | 967,83 | 996,9 | 772 | 78,95 | 98,77 | 1,94 | 1,96 | 32 | 0,0016 |
| 3 | 952,51 | 967,76 | 695 | 57,27 | 89,51 | 3,88 | 3,97 | 34 | 0,0151 |
| 2 | 929,29 | 952,01 | 720 | 38,14 | 72,84 | 5,56 | 5,84 | 29 | 0,1075 |
| 1 | 661,14 | 929,27 | 730 | 18,67 | 48,15 | 10,68 | 10,43 | 0 | 0,0478 |
| Total | | | | | | | | | 0,5032 |

Se observa que el modelo separa de forma adecuada los rangos de score con diferente morosidad, con un % Bads que se asemeja a la PD teórica en cada segmento y con un KS de 34, ampliamente superior al umbral de 20 puntos, lo que confirma la buena discriminación entre la distribución de buenos y malos pagadores, toda vez que el KS corresponde a la mayor diferencia en puntos porcentuales entre las curvas desacomuladas de goods y bads.

Gráfico I – Distribución de buenos y malos pagadores -Población de entrenamiento



3.4.3. Cálculo de Hosmer – Lemeshow:

Partiendo de la tabla scorecard realizada para la población de entrenamiento, se considera para cada rango de score el respectivo HL, que resulta de multiplicar la cantidad total de registro por el cociente entre el cuadrado de la diferencia entre la PD observada y la estimada, y el producto de la PD estimada por su complemento.

Matemáticamente:

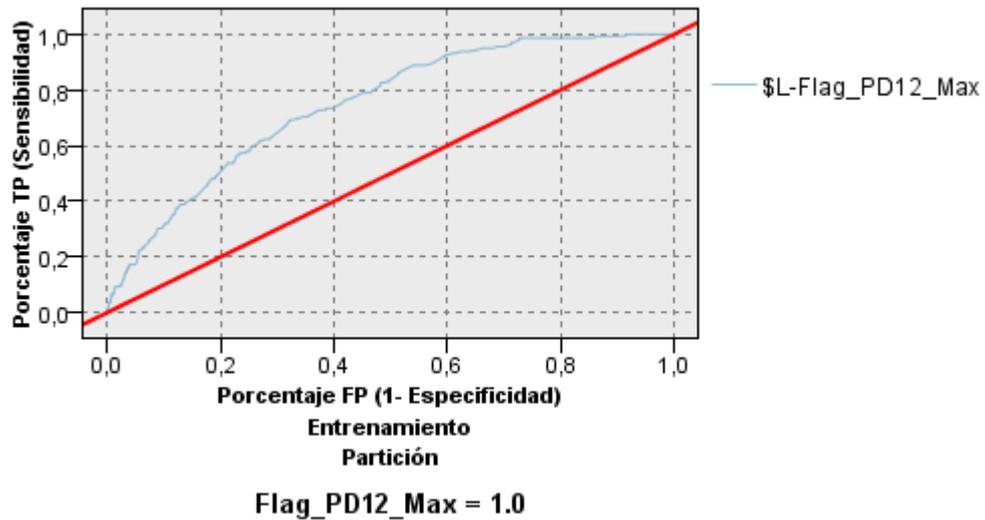
$$HL = \sum \text{Registros totales} * \frac{(PD \text{ teórica} - \% \text{ Bads})^2}{\% \text{ Bads} * (1 - \% \text{ Bads})}$$

Finalmente, el HL total resulta de la sumatoria del HL obtenido para cada rango de score. En nuestro ejemplo la sumatoria da 0,50 que al ser menor que 11,34 (valor crítico de la distribución chi cuadrado para un nivel de confianza del 1% con n-2 grados de libertad) confirma que la PD esperada se parece a la PD observada.

3.4.4. Curva ROC y área bajo la curva (AUC)

Por último, el modelo se somete al diseño de la curva ROC, que se visualiza en el gráfico adjunto, de forma de mostrar cuantos puntos de sensibilidad se pierden a medida que se ganan puntos de no especificidad.

Gráfico II – Curva ROC, población de entrenamiento



El área bajo la curva en este gráfico permite demostrar cuanto más preciso es el modelo previsto a la alternativa del azar, línea recta indiferente entre una y otra alternativa, partiendo de la base que el área bajo la curva de la recta de azar es de 50%.

Matemáticamente, el AUC se puede calcular en forma simple como una igualdad basada en el índice de Gini:

$$AUC = \frac{Gini + 1}{2}$$

Este índice se calcula creando un intervalo para cada score. En cada uno, se mide la cantidad de casos incluidos, la cantidad de buenos y la cantidad de malos. Sumarizando buenos, malos y totales se puede calcular los buenos y malos acumulados hasta el intervalo y, con ese dato, calcular el valor B y M, donde:

$$B = \frac{Tasa\ de\ buenos\ acumulados_n - Tasa\ de\ buenos\ acumulados_{n-1}}{100}$$

$$M = \frac{Tasa\ de\ malos\ acumulados_n - Tasa\ de\ malos\ acumulados_{n-1}}{100}$$

Con estos indicadores, se puede calcular el índice de Gini:

$$GINI = (1 - \sum B * M) * 100$$

Para el ejemplo, el índice arroja un valor de 0,495, lo que redonda en un área bajo la curva ROC de 74,73%, lo que confirma un poder predictivo regular, el área supera el umbral de azar a lo largo de la curva. Es preferible usar el modelo que adivinar azarosamente si el cliente será buen o mal pagador.

3.5. Testeo

Una vez realizados todas las validaciones estadísticas en la población de entrenamiento, se corre el mismo modelo en la población de validación para ver si éste mantiene las propiedades descritas anteriormente.

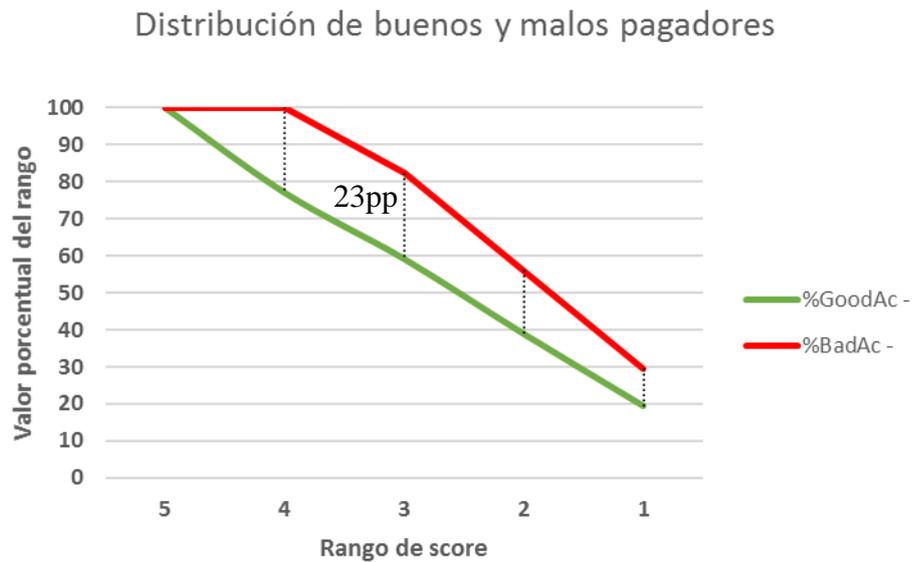
Mediante la scorecard desplegada en tabla XII, se muestra la separación en quintiles de los clientes analizados en validación y se busca observar si existe una buena discriminación entre la distribución de buenos y malos pagadores en cada rango.

Tabla XII – Scorecard población validación.

| <i>Rango Score</i> | <i>Score Mín.</i> | <i>Score Máx.</i> | <i>#Total</i> | <i>%Good Ac -</i> | <i>%Bad Ac -</i> | <i>%Bads</i> | <i>PD Teórica</i> | <i>KS</i> | <i>HL</i> |
|--------------------|-------------------|-------------------|---------------|-------------------|------------------|--------------|-------------------|-----------|---------------|
| 5 | 997 | 999,75 | 205 | 100 | 100 | 0 | 0,16 | 22 | 0,0000 |
| 4 | 970,05 | 996,92 | 166 | 77,02 | 100 | 3,61 | 1,7 | 23 | 1,7403 |
| 3 | 954,87 | 969,9 | 190 | 59,08 | 82,35 | 4,74 | 3,76 | 17 | 0,4041 |
| 2 | 930,66 | 954,77 | 183 | 38,79 | 55,88 | 4,92 | 5,62 | 10 | 0,1917 |
| 1 | 661,14 | 929,76 | 182 | 19,28 | 29,41 | 5,49 | 10,14 | 0 | 7,5845 |
| Total | | | | | | | | | 9,9206 |

Se observa que la validación del modelo separa de forma adecuada los rangos de score con diferente morosidad, con un KS de 23, que, aunque más bajo que el observado en entrenamiento, sigue siendo superior al umbral de 20 puntos, lo que confirma la buena discriminación entre la distribución de buenos y malos pagadores.

Gráfico III – Distribución de buenos y malos pagadores -Población de validación



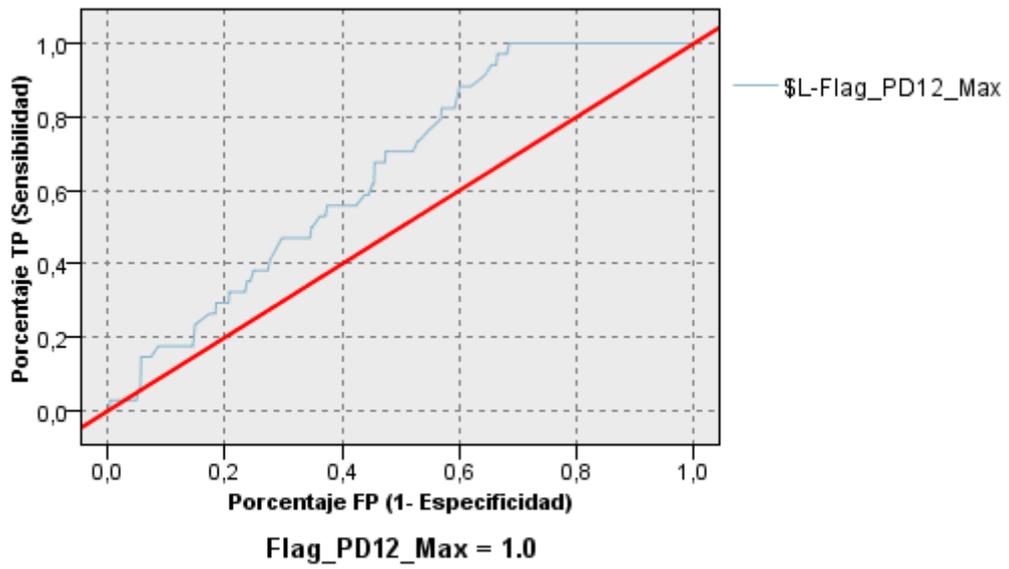
En cuanto al estadístico de Hosmer – Lemeshow, el mismo indica que, en validación, la PD teórica ajusta peor a la PD empírica que en entrenamiento, pero con un valor de sumatoria de 9,92. Al ser menor que 11,34 (valor crítico de la distribución chi cuadrado para un nivel de confianza del 1% con n-2 grados de libertad) confirma que la PD esperada se parece a la PD observada, lo que también se puede apreciar al comparar las columnas de %Bads y PD teórica.

Por último, y en forma consistente al análisis, se muestra en la tabla XIII una curva ROC más aplanada que la provista en el entrenamiento, pero también más eficiente que el azar, en tanto los resultados obtenidos para el índice de Gini y el área bajo la curva fueron de 30,35 y 65,18% respectivamente.

Tabla XIII – KS, GINI y AUC - Comparativo poblaciones

| Entrenamiento | | | Validación | | |
|---------------|-------|-------|------------|-------|-------|
| KS_Max | GINI | AUC | KS_Max | GINI | AUC |
| 34 | 49,46 | 74,73 | 23 | 30,35 | 65,18 |

Gráfico IV – Curva ROC, población de entrenamiento



Capítulo IV Modelización digital

En el capítulo anterior se procedió a evaluar un modelo tradicional de riesgo de crédito aplicado a la base de datos utilizada en esta tesis. Una vez corroborada la estabilidad, poder predictivo y alcance del modelo tradicional se procede a la incorporación de variables digitales. Se busca de esta manera comparar el nuevo modelo con el tradicional para validar o refutar la hipótesis rectora de este trabajo, consistente en proponer un modelo de evaluación crediticia que sea aplicable al mercado de créditos de consumo e incluya variables digitales de e-mail y social media, para ser utilizados en un entorno *Fintech*, considerando el período 2016-2018.

La modelización digital toma como punto de partida todo el set de variables analizado en el capítulo anterior, tanto las utilizadas como las descartadas por el modelo, e incorpora un set adicional de variables explicativas del comportamiento de los clientes en el mercado de crédito provenientes del entorno *FinTech*. El detalle de su tipología puede encontrarse en la tabla I del anexo IV, mientras que los principales indicadores de estadística descriptiva se desarrollan en la tabla II del mismo apartado.

4.1. Variables digitales incorporadas

Estas nuevas variables provienen de tres fuentes primigenias de información: Las que surgen del e-mail declarado por el cliente, las asociadas a su teléfono de línea o celular y las referentes a su cuenta de Twitter asociada al mail declarado. Todas ellas tienen la particularidad de poder esconder o reflejar datos reales del propietario a conveniencia. Por ejemplo, una persona puede incorporar en su dirección de mail su nombre y apellido o apelar a *nicknames* o incluso texto ficcional; la cuenta de Twitter puede asociarse a un mail personal o a una casilla anónima y un teléfono se puede vincular a un domicilio real si se trata de uno de línea, pero uno celular, móvil por definición, permite múltiples usuarios y no cuenta con un directorio que lo asocie con su propietario.

4.1.1. Variables asociadas al e-mail

Para el caso de la información que se desprende del e-mail declarado por el cliente, se construyó un nuevo set de variables dicotómicas. Al tratarse de una cadena de texto, puede compararse con otros fragmentos de información personal e inferir una serie de características distintivas de los usuarios. De este modo, **“nombre en mail”** y **“Apellido en**

mail” verifica que el nombre o el apellido del cliente se vean reflejados en su dirección de mail, lo que podría asociarse con una mayor predisposición de cada individuo a exponerse en forma real. Para el universo analizado, el 44% de los casos tienen su nombre reflejado en el mail, valor que asciende al 48% al tratarse de su apellido. Las variables **“Guiones en mail”** y **“Punto en mail”** trabajan en sentido similar, buscando apariciones de caracteres especiales, el guion y el punto, dentro del *username* del mail. La cantidad de apariciones en estos casos se reduce drásticamente, alcanzando el 11% en el caso de los guiones y 8% para los puntos.

Dado que la dirección de mail es declarativa en la solicitud de crédito y el cliente pudo haber entregado un mail “basura”, se procedió a una verificación manual del *string* del cuerpo del mail, buscando que no asuma un comportamiento irracional. Por Ej: “aaaa@abc.com”. Estos resultados se reflejaron en la variable *dummy* **“Es mail basura?”** con 1 para los afirmativos y 0 para los negativos. En similar sentido, se contabilizó la cantidad de veces que aparecía cada dirección de mail y, debido a que se trabajó con cada cliente una única vez, se registraron aquellos correos que aparecían en más de una oportunidad bajo la variable dicotómica **“Email repetido”**. Ninguna de las dos entrega guarismos significativos.

En cuanto al lado derecho de la dirección de mail, llamada dominio, se puede deducir cual es el proveedor de servicio de mail, resumido en un set de variables dummies llamadas **“Es gmail”, “Es outlook”, “Es hotmail”** y **“Es yahoo”**. Si todas toman valor 0, se trata de otro servidor. Casi el 60% de los casos corresponde al servidor Gmail y sumado al 31% de Hotmail predominan en el universo analizado.

Así mismo, por la extensión del dominio se puede saber si se trata de un servicio local o extranjero. Por ejemplo: .ar refiere a dominios radicados bajo jurisdicción nacional, dato consolidado en la variable dicotómica **“Origen servidor”**, la que otorga una tasa de aparición del 91,46% para servidores extranjeros.

4.1.2. Variables asociadas al número telefónico

Así como el correo electrónico es un dato de tipo declarativo, el número telefónico de contacto se presenta como una cadena numérica que puede ser analizada o comparada contra la información que el bureau externo de crédito tiene sobre los últimos teléfonos declarados por el individuo. Si este dato es consistente con lo declarado en la solicitud, podemos inferir que se trata de un número real. Mediante la variable **“DeclaraCel”**, la que

toma valor 1 cuando un cliente declara un teléfono que comienza con el prefijo 15, se verifica que el solo el 15% de la muestra facilita un celular como teléfono de contacto. El 85% restante, que declara un teléfono de línea, presupone una muy buena tasa de contactabilidad con el cliente. Sin embargo, solo el 61% de los clientes declara un teléfono conocido por el bureau, lo que mitiga esta premisa. El dato se refleja en la variable dummy **“CoincideTelDec”**.

4.1.3. Variables asociadas a la cuenta de Twitter

Tal como se mencionara previamente, el acceso a la red social Twitter puede hacerse a través de una cuenta de correo personal o bien con una específica a tal efecto, por lo que la asociación de un mismo e-mail para solicitar un préstamo y participar de una red social podría indicarnos una cierta consistencia en la información, así como inferir particularidades del comportamiento de sus usuarios, de una red que se presume libre y anónima.

Mediante la variable **“Tiene twitter asociado”** se capturaron aquellos casos en los que el mail declarado por el cliente se encuentra registrado en esa red social, independientemente a que cuenta se encuentre asociado. El 24,19% de la base cumplía con este requisito.

Por último, dado que Twitter permite saber si la cuenta registrada cuenta con un teléfono móvil asociado, se buscaron coincidencias entre las últimas dos cifras de teléfono declaradas en esa red social con los números informados por el bureau, de forma de inferir si el cliente utilizaba el mismo teléfono para todas sus operaciones. **“Coincide cel”** fue la variable creada a tal fin, con apenas un 1% de casos validados, lo que reduce el rigor estadístico de la misma.

4.2. Análisis bivariado

Del mismo modo que en el capítulo anterior, se realiza un análisis bivariado de las variables recientemente presentadas sobre la variable objetivo, es decir, el promedio de clientes que hayan caído en default en los últimos doce meses ($\text{flag_PD12} = 1$) para cada una de ellas.

Dentro de las asociadas al e-mail y centrados en el *username*, la variable **“Apellido en mail”** ayuda a discriminar mejor entre buenos y malos pagadores, encontrándose una diferencia de 20% en favor de quienes declaran su apellido en el mail a quienes no lo hacen. Sin embargo, el **“nombre en mail”**, con solo 0.12pp de contraste, no permite discernir con

facilidad entre unos y otros. Con esto se podría inferir que una persona que asocia su apellido familiar al medio de contacto tiene mayor predisposición al pago que quienes no lo hacen.

Para la identificación de puntos o guiones, se encuentra escasa evidencia de que puedan aportar a discriminar buenos o malos pagadores. La variable **“Punto en mail”** no muestra materialidad y **“Guiones en mail”** muestra apenas un 13% de aumento en la tasa de *default* de clientes que utilizan el signo “-” o “_” en su casilla de correos.

Analizando las variables desde el lado del dominio, se observa que la cantidad de repeticiones o la calidad del mail informado no entrega materialidad para diferenciar comportamientos, al igual que la mayoría de las características asociadas al servidor de correo electrónico. Aun así, se destaca que quienes declaran una cuenta de Gmail alcanzan una tasa de default del 4,55%, contra un 3,89% de quienes eligen otra alternativa.

Focalizar sobre el número telefónico demuestra consistencia a la hora de medir comportamiento crediticio. Por un lado, se verifica que la variable **“DeclaraCel”** muestra mejor comportamiento entre quienes declaran un teléfono fijo, aunque con apenas un 10% de diferenciación entre ambas alternativas. Por su parte, **“CoincideTelDec”** se muestra más robusta en el mismo sentido, al inferir que aquellos clientes que declaren un número telefónico conocido por el bureau de crédito tienen una tasa de default más del 30% menor.

Por último, la variable **“Tiene twitter asociado”** muestra que quienes asocian su mail con su cuenta de Twitter son peores pagadores, a razón de casi un 20% por sobre los que no se encuentran vinculados, sin poder conseguir información sólida respecto de la variable **“Coincide cel”** en esta red social.

4.3. Modelo digital

Luego del análisis bivariado de las variables digitales, se las incorpora al set de variables originalmente consideradas para realizar el modelo tradicional con nuevas características. Del mismo modo que en el capítulo anterior, se procedió a renombrar las variables a considerar con el sufijo “_S”, de forma de indicar que ya no se trata de un dato en crudo, sino de uno tamizado.

Utilizando los mismos grupos generados en la partición de entrenamiento y validación y siguiendo los mismos parámetros metodológicos descriptos anteriormente, se construyó un nuevo modelo que considere todas las variables disponibles, independientemente de que se hayan utilizado o no para construir el modelo tradicional,

repetiendo el objetivo de predecir la variable independiente “*Flag_pd_12*” a través de una serie de variables explicativas de este fenómeno.

De las condiciones antes explicadas se desprende un modelo cuyas variables incorporadas fueron “*Edad_S*”, “*Actividad 1_S*”, “*Region_S*”, “*Tipo de pago_S*”, “*PSF13_S*”, “*Capital solicitado_S*”, “*Bancos 12m_S*”, “*Sit BCRA 24m_S*”, “*Exposicion 24m BCRA_S*”, “*Apellido en mail_S*”, “*Tiene twitter asociado_S*” y “*CoincideTelDec_S*”. Entonces, la nueva probabilidad de default aplicable a cada cliente responde a la siguiente ecuación:

$$PD = \frac{1}{1 + e^{(1,548+0,326*Edad_S(1)+0,616*Edad_S(2)+0,466*Actividad\ 1_S(1)+0,370*Region_S(1)+3,791*Tipo\ de\ pago_S(1)-0,512*PSF13_S(1)-0,394*Capital\ solicitado_S(1)-0,697*Capital\ solicitado_S(2)+0,627*Bancos\ 12m_S(1)-0,603*Sit\ BCRA\ 24m_S(1)+0,935*Exposicion\ 24m\ BCRA_S(1)+0,701*Exposicion\ 24m\ BCRA_S(2)+0,517*Exposicion\ 24m\ BCRA_S(3)+0,374*Apellido\ en\ mail_S(1)-0,169*Tiene\ twitter\ asociado_S(1)+0,500*CoincideTelDec_S(1))}}$$

De esta fórmula se desprende el Score, entendido como $(1 - PD) * 1000$.

Su impacto medido en términos de β , los grados de libertad asociados a la cantidad de instancias de cada una (expresados como n-1) y su prueba de significancia medida a través del test de Wald se detallan en la tabla XIV.

Tabla XIV - Resultado del modelo

| Variables | B | Error estándar | Wald | gl | Sig. |
|------------------------------------|--------|----------------|--------|----|------|
| <i>Edad_S</i> | | | 4,991 | 2 | ,082 |
| <i>Edad_S(1)</i> | -,326 | ,189 | 2,988 | 1 | ,084 |
| <i>Edad_S(2)</i> | -,616 | ,304 | 4,111 | 1 | ,043 |
| <i>Actividad 1_S(1)</i> | -,466 | ,199 | 5,466 | 1 | ,019 |
| <i>Region_S(1)</i> | -,370 | ,171 | 4,705 | 1 | ,030 |
| <i>Tipo de pago_S(1)</i> | -3,791 | ,750 | 25,542 | 1 | ,000 |
| <i>PSF13_S(1)</i> | ,512 | ,237 | 4,684 | 1 | ,030 |
| <i>Capital solicitado_S</i> | | | 9,512 | 2 | ,009 |
| <i>Capital solicitado_S(1)</i> | ,394 | ,237 | 2,764 | 1 | ,096 |
| <i>Capital solicitado_S(2)</i> | ,697 | ,229 | 9,299 | 1 | ,002 |
| <i>Bancos 12m_S(1)</i> | -,627 | ,248 | 6,374 | 1 | ,012 |
| <i>Sit BCRA 24m_S(1)</i> | ,603 | ,184 | 10,721 | 1 | ,001 |
| <i>Exposicion 24m BCRA_S</i> | | | 8,947 | 3 | ,030 |
| <i>Exposicion 24m BCRA_S(1)</i> | -,935 | ,363 | 6,628 | 1 | ,010 |
| <i>Exposicion 24m BCRA_S(2)</i> | -,701 | ,398 | 3,110 | 1 | ,078 |
| <i>Exposicion 24m BCRA_S(3)</i> | -,517 | ,413 | 1,570 | 1 | ,210 |
| <i>Apellido en mail_S(1)</i> | -,374 | ,170 | 4,826 | 1 | ,028 |
| <i>Tiene twitter asociado_S(1)</i> | ,169 | ,186 | ,828 | 1 | ,363 |
| <i>CoincideTelDec_S(1)</i> | -,500 | ,171 | 8,560 | 1 | ,003 |
| <i>Constante</i> | -1,548 | ,419 | 13,658 | 1 | ,000 |

Los resultados de β permiten inferir, nuevamente, cuál será la relación entre poseer determinada característica y la posibilidad de *default*, esto sujeto a la codificación de las variables provista por el software estadístico, que se detalla en tabla XV. Dado que los resultados obtenidos son consistentes con el modelo tradicional, solo se comentan las nuevas variables agregadas.

El nuevo set incluye dos características dejadas de lado por el anterior. “*Capital solicitado*”, con relación directa entre la cantidad de dinero que se pide prestada y la probabilidad de impago, lo que muestra un total sentido económico y “*Exposicion 24m BCRA_S*”, cuya relación inversa lleva a pensar que, a menor cantidad de deuda contraída en los últimos dos años, mayor probabilidad de default.

Dentro de las variables tomadas del entorno digital, se observa que quienes tienen su apellido incorporado a su cuenta de mail son mejores pagadores, pero quienes tienen Twitter asociado a ese correo son peores pagadores. Adicionalmente, aquellos cuyo teléfono declarado fue detectado por el bureau presentan, lógicamente, una mayor voluntad de pago que los que evitan dar datos conocidos.

La suma de esta información permite inferir que el hecho de proporcionar datos ciertos, verificables y asociados al cliente se condicen con individuos más propensos a cumplir con sus obligaciones crediticias.

Tabla XV - Codificaciones de variables categóricas agregadas

| Variable | R | Frec | Codific param | | |
|------------------------------|-----|------|---------------|----|----|
| | | | -1 | -2 | -3 |
| <i>Exposicion 24m BCRA_S</i> | 1.0 | 413 | 0 | 0 | 0 |
| | 2.0 | 1255 | 1 | 0 | 0 |
| | 3.0 | 1054 | 0 | 1 | 0 |
| | 4.0 | 932 | 0 | 0 | 1 |
| <i>Capital solicitado_S</i> | 1.0 | 901 | 0 | 0 | |
| | 2.0 | 1181 | 1 | 0 | |
| | 3.0 | 1572 | 0 | 1 | |

| Variable | R | Frec | -1 |
|---------------------------------|-----|------|----|
| | | | |
| <i>Tiene twitter asociado_S</i> | 0.0 | 2758 | 0 |
| | 1.0 | 896 | 1 |
| <i>Apellido en mail_S</i> | 0.0 | 1886 | 0 |
| | 1.0 | 1768 | 1 |
| <i>CoincideTelDec_S</i> | 0.0 | 1414 | 0 |
| | 1.0 | 2240 | 1 |
| | | | |

Tal como se explicara en el capítulo anterior, algunas variables se mantuvieron en el nuevo modelo a pesar de no ser estadísticamente significativas ($\text{sig} > 10\%$). En este caso se trata de la variable “*Tiene twitter asociado_S*”, la que se mantiene para realizar una primera vinculación entre la participación en esta red social y la probabilidad de default de un cliente.

En forma complementaria y consistente con lo anteriormente descrito, se refleja en la tabla XVI el poder predictivo de cada una de las nuevas variables medido en forma individual a través de sus respectivos KS, GINI y AUC, verificándose que cada una de ellas tiene un área bajo la curva mayor al 50%, por lo que las hace preferente al azar.

Tabla XVI – Poder predictivo individual

| Variable | KS | GINI | AUC |
|---------------------------------|----|------|-------|
| <i>Tiene twitter asociado_S</i> | 4 | 4,05 | 52,03 |
| <i>Apellido en mail_S</i> | 7 | 7,35 | 53,68 |

| Variable | KS | GINI | AUC |
|-------------------------|----|------|-------|
| <i>CoincideTelDec_S</i> | 9 | 9,24 | 54,62 |

4.4. Pruebas de estimador

Una vez determinado el nuevo modelo predictor de la probabilidad de impago, se debe probar que tenga relevancia estadística con las mismas herramientas utilizadas *ut supra*.

4.4.1. Matriz de clasificación

Las tablas XVII y XVIII muestran que clientes se presumen que caerán en *default* utilizando nuevamente la tasa de corte de 0,059 y asignando 1 a quienes superen ese rango.

Tabla XVII - Tabla de clasificación

| Observado | | | Pronosticado | | |
|-------------------|---------------|-----|---------------|-----|---------------------|
| | | | Flag_PD12_Max | | Porcentaje correcto |
| | | | 0.0 | 1.0 | |
| Paso 1 | Flag_PD12_Max | 0.0 | 2593 | 899 | 74,3 |
| | | 1.0 | 60 | 102 | 63,0 |
| Porcentaje global | | | 3654 | | 73,8 |

Tabla XVIII - Matriz de clasificación (en %)

| Observado / Pronosticado | Positivo | Negativos | Totales |
|--------------------------|----------|-----------|---------|
| Positivos | 2,80% | 1,64% | 4,44% |
| Negativos | 24,60% | 70,96% | 95,56% |
| Totales | 27,88% | 72,12% | 100,00% |

Comparando estas tablas con las obtenidas con su acápite análogo, se desprenden conclusiones relevantes, tanto desde una visión negocio como desde el punto de vista de la inclusión financiera. A nivel de quien presta el dinero, el nuevo modelo predice mayor cantidad de morosos con un 4,08% más de eficacia, lo que puede ayudar a bajar las previsiones de incumplimiento, reduciendo el capital inmovilizado. Visto con perspectiva de consumidor, predice asimismo mayor cantidad de buenos pagadores en términos de 1 por cada mil solicitantes, lo que permite en este ejemplo incorporar al mercado de crédito a 22 personas que hubieran sido rechazadas por un modelo tradicional y, por efecto del menor riesgo sistémico, reducir la tasa de interés a la que se pacta el préstamo. El resumen comparativo se detalla en la tabla XIX.

Tabla XIX - Tabla de clasificación comparativa

| Observado | | | Pronosticado | | |
|-------------------|---------------|------|---------------|--------------|---------------------|
| | | | Flag_PD12_Max | | Porcentaje correcto |
| | | | 0.0 | 1.0 | |
| Paso 1 | Flag_PD12_Max | 0.0 | +22 (+0,08%) | -22 (-2,39%) | +0,7 |
| | | 1.0 | -4 (-6.25%) | +4 (+4,08%) | +2,5 |
| Porcentaje global | | 3654 | | | +0,8% |

4.4.2. Scorecard

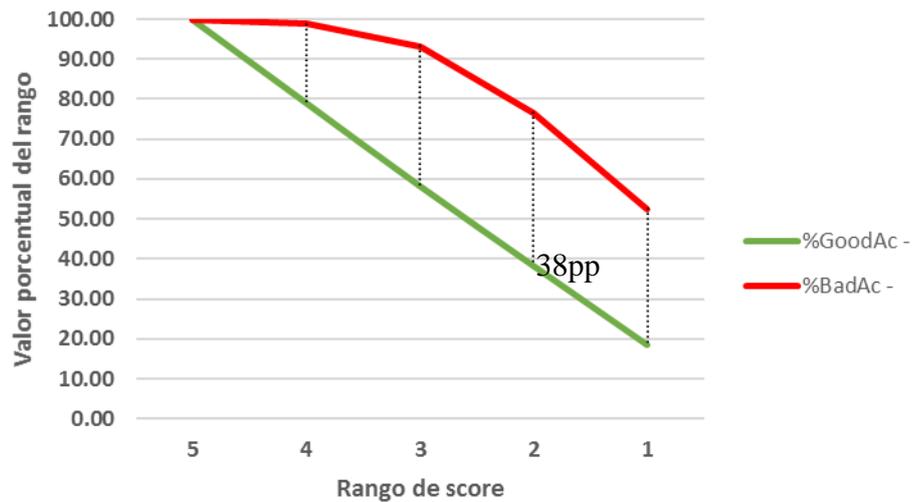
A los fines de corroborar si existe una buena discriminación entre la distribución de buenos y malos pagadores en cada rango se procede a realizar la *scorecard* del nuevo modelo, detallada en tabla XX.

Tabla XX – Scorecard población entrenamiento

| Rango Score | Score Mín. | Score Máx. | #Total | %Good Ac - | %Bad Ac - | %Bads | PD Teórica | KS | HL |
|--------------|------------|------------|--------|------------|-----------|-------|------------|----|---------------|
| 5 | 996,56 | 999,84 | 730 | 100,00 | 100,00 | 0,27 | 0,14 | 19 | 0,4581 |
| 4 | 972,49 | 996,56 | 736 | 79,15 | 98,77 | 1,22 | 1,59 | 34 | 0,8360 |
| 3 | 955,63 | 972,47 | 726 | 58,33 | 93,21 | 3,72 | 3,57 | 38 | 0,0456 |
| 2 | 928,06 | 955,60 | 731 | 38,32 | 76,54 | 5,34 | 5,63 | 33 | 0,1216 |
| 1 | 616,03 | 928,02 | 731 | 18,50 | 52,47 | 11,63 | 11,25 | 0 | 0,1027 |
| Total | | | | | | | | | 1,5642 |

Se verifica que el modelo separa de forma adecuada los rangos de score con diferente morosidad, con un % Bads que se asemeja a la PD teórica en cada segmento y con un KS de 38, un 11,73% mayor que en el modelo tradicional y muy superior al umbral de 20 puntos, lo que confirma una mejor discriminación entre buenos y malos pagadores. Gráficamente:

Gráfico V – Distribución de buenos y malos pagadores -Población de entrenamiento



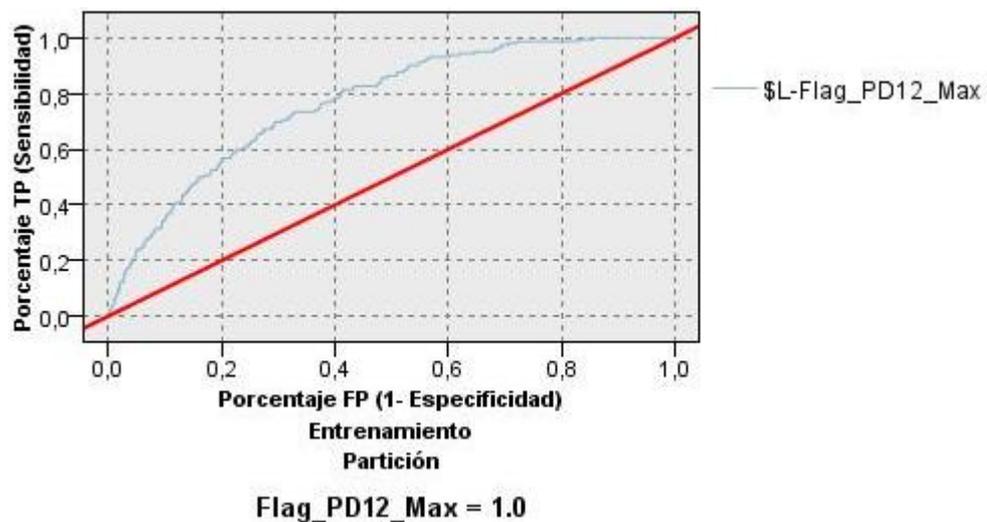
4.4.3. Cálculo de Hosmer – Lemeshow:

El HL, sumatoria del HL obtenido para cada rango de score, ofrece un valor de 1,56 que sigue siendo menor que el valor crítico de la distribución chi cuadrado 11,34 utilizado como umbral, lo que confirma que la PD esperada se parece a la PD observada.

4.4.4. Curva ROC y área bajo la curva (AUC)

Por último, el modelo se somete al diseño de la curva ROC, que se visualiza en el gráfico VI, de forma de mostrar cuantos puntos de sensibilidad se pierden a medida que se ganan puntos de no especificidad.

Gráfico VI – Curva ROC, población de entrenamiento



Dado que la sumatoria de buenos y malos pagadores ofrece un índice de Gini con valor de 0,539, lo que redunda en un área bajo la curva ROC de 76,92%, se confirma un

poder predictivo superior al modelo anterior en 2,19pp, lo que hace preferible usar esta propuesta para determinar si el cliente será buen o mal pagador.

4.5. Testeo

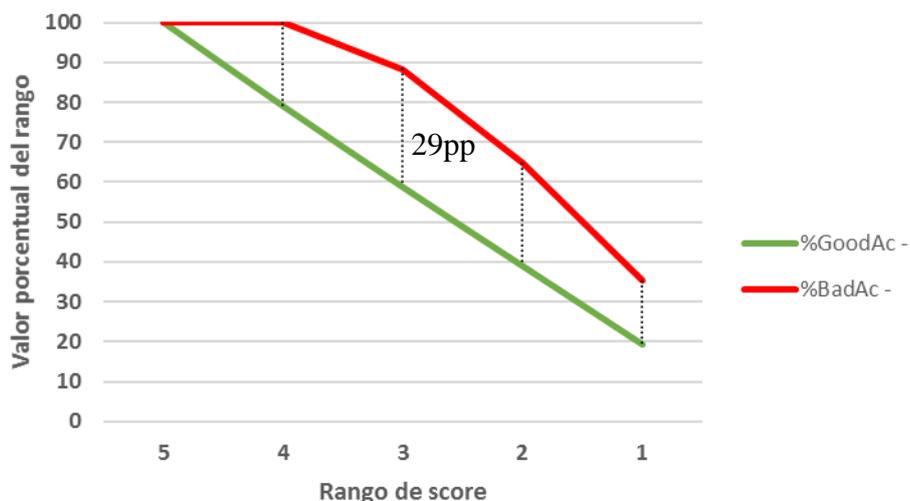
Tal como se realizara en el capítulo III, se procede a correr el nuevo modelo en la población de validación, para confirmar si se mantienen las propiedades descritas en el grupo de entrenamiento. Con la scorecard de tabla XXI se muestra la separación en quintiles de los clientes analizados en validación para determinar si existe una buena discriminación entre la distribución de buenos y malos pagadores en cada rango.

Tabla XXI – Scorecard población validación.

| Rango Score | Score Mín. | Score Máx. | #Total | %Good Ac - | %Bad Ac - | %Bads | PD Teórica | KS | HL |
|--------------------|-------------------|-------------------|---------------|-------------------|------------------|--------------|-------------------|-----------|---------------|
| 5 | 997,36 | 999,73 | 187 | 100 | 100 | 0 | 0,13 | 20 | 0 |
| 4 | 975,11 | 997,35 | 185 | 79,04 | 100 | 2,16 | 1,34 | 29 | 0,5886 |
| 3 | 955,52 | 975,05 | 184 | 58,74 | 88,24 | 4,35 | 3,44 | 25 | 0,3662 |
| 2 | 932,9 | 955,39 | 185 | 39,01 | 64,71 | 5,41 | 5,44 | 15 | 0,0003 |
| 1 | 549,98 | 932,52 | 185 | 19,39 | 35,29 | 6,49 | 10,85 | 0 | 5,7948 |
| Total | | | | | | | | | 6,7499 |

La validación del nuevo modelo separa de forma adecuada los rangos de score con diferente morosidad, con un KS de 29, un 26% superior al observado en el tradicional y al umbral de 20 puntos, lo que reconfirma la buena discriminación de pagadores. Gráficamente:

Gráfico V – Distribución de buenos y malos pagadores -Población de validación



El estadístico de Hosmer – Lemeshow indica que, en validación, la PD teórica ajusta peor a la PD empírica que en entrenamiento, pero con un valor de sumatoria de 6,74 que al

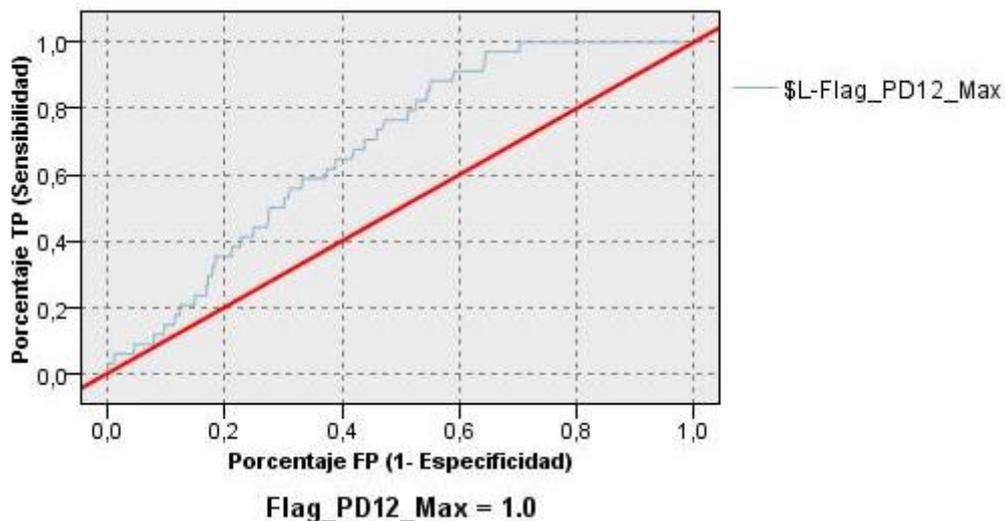
ser menor que 11,34 (valor crítico de la distribución chi cuadrado) confirma nuevamente que la PD esperada se parece a la PD observada, lo que también se puede apreciar al comparar las columnas de %Bads y PD teórica.

Por último, se muestra en el gráfico VI una curva ROC más aplanada que la provista en el entrenamiento, pero también más eficiente que el azar, en tanto los resultados obtenidos para el índice de Gini y el área bajo la curva fueron de 36,62 y 68,31% respectivamente, según se observa la tabla XXII.

Tabla XXII – KS, GINI y AUC - Comparativo poblaciones

| Entrenamiento | | | Validación | | |
|---------------|-------|-------|------------|-------|-------|
| KS_Max | GINI | AUC | KS_Max | GINI | AUC |
| 38 | 53,84 | 76,92 | 29 | 36,62 | 68,31 |

Gráfico VI – Curva ROC, población de entrenamiento



4.6. Consideraciones

La comparativa entre el modelo propuesto en este apartado y uno de tipo tradicional como el que se describe en el capítulo anterior permite forjar algunos conceptos que serán fundamentales al momento de propiciar el uso de variables del entorno *FinTech* en modelos predictivos, ya sea como una mejora a nivel negocio crediticio como desde la inclusión financiera. Este nuevo modelo no solo predice mayor cantidad de morosos en forma más eficaz, sino que también incrementa el conjunto de buenos pagadores discernidos.

Dado que los inputs del modelo *FinTech* se nutren de información digitalizada y de fácil acceso, incluir estas variables digitales tiene un costo extremadamente bajo para los

consumidores del modelo y permite mejorar su calibración incluyendo más personas, las que a su vez robustecen el sistema con nuevos inputs en un círculo virtuoso. En este caso particular, incorporando a 22 personas que hubieran sido rechazadas por un modelo tradicional.

Capítulo V Conclusiones

A lo largo de este trabajo se buscó demostrar la importancia de sumar variables del entorno *FinTech* en modelos predictivos de riesgo, con el fin de proponer un modelo de evaluación crediticia que sea aplicable al mercado de créditos de consumo e incluya variables digitales de e-mail y social media, analizando una variable de comportamiento de pago, o variable respuesta, que determine si el cliente resultó malo o bueno en términos de cumplimiento durante los 12 meses siguientes a la fecha de observación.

Una primera etapa consistió en el armado de un modelo tradicional de scoring que sirva como punto comparativo o “*baseline*” y, gracias a esto, identificar y caracterizar variables que se consideren actualmente relevantes para la evaluación crediticia online en el segmento de créditos de consumo. Luego, además de mostrar cómo se conforma un modelo tradicional, se incorporaron y testearon nuevas variables de correo electrónico, teléfono fijo, celular y Twitter que pudieran ser incorporadas a futuros modelos de scoring crediticio.

La comparativa entre el modelo tradicional propuesto en el capítulo tercero y uno de tipo digital como el que se describe en el capítulo anterior exhibe diversas ventajas que se resumen a continuación:

- **Aumento del poder predictivo en morosidad:** Predice mayor cantidad de morosos con una eficacia superior al 4%, que se traduce en argumentos concretos para reducir las provisiones por incumplimiento por parte de la industria.
- **Aumento del poder predictivo en pagadores:** Incrementa el conjunto de buenos pagadores discernidos en alrededor de 1 por cada mil solicitantes respecto al modelo tradicional, lo que redundará en una reducción de la tasa de interés teórica por efecto del menor riesgo sistémico.
- **Información a bajo costo:** Por la naturaleza digital y de fácil acceso de los inputs del modelo *FinTech*, incluir estas variables digitales en un modelo predictivo tiene un costo extremadamente bajo.
- **Retroalimentación:** Derivado de los puntos anteriores, la incorporación de más aprobados permite mejorar la calibración del modelo resultante, los que a su vez robustecen el sistema con nuevos inputs en un círculo virtuoso.

Identificar buenos pagadores que antes se hubieran rechazado ayuda a la inclusión financiera al incorporar potenciales clientes que no hubieran sido aceptados. Dicho de otra manera, fueron los datos relacionados con el entorno *FinTech* los que permitieron que un segmento de clientes que no hubieran pasado un “score” tradicional sean aprobados y sumados a un ecosistema financiero que de otra manera les hubiera dado la espalda.

Esto prueba la hipótesis que dio origen a este trabajo de investigación. Queda comprobado que la incorporación de variables digitales como tener el apellido inscrito en el dominio del e-mail o asociar ese correo a un perfil social como Twitter mejora los modelos tradicionales y predice de forma más eficiente la probabilidad de pago de créditos de consumo permitiendo incorporar nuevos tomadores de créditos que actualmente se encuentran excluidos por falta de información.

A modo de cierre se plantea una serie de recomendaciones y pasos a seguir, derivados de la investigación realizada:

1. **Dotar al modelo de una base mayor para probar la eficiencia de las variables digitales:** En este informe se trabajó con un *dataset* de unos 4500 clientes de un banco en particular, por lo que mayor cantidad de individuos de naturaleza más diversa ayudaría a aplicar el modelo propuesto en otras poblaciones.
2. **Indagar nuevas fuentes de información en el entorno *FinTech*:** Existen diversos canales para la obtención de datos en el universo tecnológico que no han sido abarcados en este reporte. Queda mucho por indagar en el campo de herramientas de minería de texto y ponderación de comportamientos (*like*, *dislike*) de los que se basan los principales decisores de comportamiento de la industria del conocimiento. También para revalidar características del estilo “perfil social actualizado” o “cantidad de seguidores” propuestas en la hipótesis rectora.
3. **Proponer un nuevo modelo dotado exclusivamente de variables *FinTech*:** En el extremo, se entiende como “población débil” a los individuos que no cuentan con ningún tipo de información de comportamiento crediticio. Si bien el modelo digital aquí propuesto pondera variables digitales, se alimenta también de la información crediticia provista por los bureaus. La propuesta para un futuro trabajo consiste en diseñar un modelo que consiga predecir la probabilidad de default utilizando solamente variables digitales. De esta

forma se podría avanzar aún más en el objetivo de alcanzar una mayor inclusión financiera e igualdad de oportunidades para todos.

Referencias bibliográficas

- Accenture. (2018). *Informe ecosistema Fintech Argentino*. Recuperado de <https://www.afluente.com/Informe-Fintech-Argentina-2018>
- Altman, E. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". *Journal of Finance*. 23 (4): 589–609. doi:10.1111/j.1540-6261.1968.tb00843.x.
- Banco Mundial (2018). *Inclusión financiera*. Recuperado de <https://www.bancomundial.org/es/topic/financialeconomicinclusion/overview#1>
- Banco Central de la República Argentina [BCRA]. (2019). *Central de información. Texto ordenado*. Recuperado de http://www.bcra.gov.ar/SistemasFinancierosYdePagos/Ordenamiento_y_resumenes.asp
- Banco Central de la República Argentina [BCRA]. (2019). *Clasificación de deudores. Texto ordenado*. Recuperado de http://www.bcra.gov.ar/SistemasFinancierosYdePagos/Ordenamiento_y_resumenes.asp
- Banco Central de la República Argentina [BCRA]. (2019). *Cuadros estandarizados de series estadísticas*. Recuperado de https://www.bcra.gob.ar/PublicacionesEstadisticas/Cuadros_estandarizados_series_estadisticas.asp
- Banco Central de la República Argentina [BCRA]. (2019). *Gestión crediticia. Texto ordenado*. Recuperado de http://www.bcra.gov.ar/SistemasFinancierosYdePagos/Ordenamiento_y_resumenes.asp

- Banco Central de la República Argentina [BCRA]. (2018). *Política financiera*. Recuperado de http://www.bcra.gov.ar/SistemasFinancierosYdePagos/Politica_Financiera.asp
- BBVA. (2012). *Metodologías de cuantificación del riesgo de crédito*. Recuperado de <https://accionistaseinversores.bbva.com/microsites/bbva2012/es/Gestiondelriesgo/ Metodologiasdecuantificaciondelriesgodecredito.html>
- Bessis, J. (2002). *Risk management in banking*. West Sussex, Inglaterra: World Bank: John Wiley & Sons Ltd.
- Breiman, L., Friedman, J., Olsen, R., Stone, C. (1984). *Classification and regression trees*. Washington DC, Estados Unidos: World Bank: Chapman & Hall
- Boyd, D. y Ellison, N. (2007). *Social Network Sites: Definition, History, and Scholarship*. doi:10.1111/j.1083-6101.2007.00393.x
- Comité de Basilea. (2000). *Principles for the Management of Credit Risk* (1). Recuperado de <https://www.bis.org/publ/bcbs75.pdf>
- Comité de Basilea. (2004). *Convergencia internacional de medidas y normas de capital* (1). Recuperado de <https://www.bis.org/publ/bcbs107esp.pdf>
- Comité de Basilea. (2016). *Prudential treatment of problem assets - definitions of non-performing exposures and forbearance* (1). Recuperado de <https://www.bis.org/publ/d367.pdf>
- Demirgüç-Kunt, A., Klapper, L., Singer, D., Saniya, A. y Hess, J (2018). *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. Washington DC, Estados Unidos: World Bank.
- Ernst & Young. (2017). *Guía EY sobre Fintech, una evaluación sobre Fintech en el Reino Unido y a nivel global*. Reino Unido: Ernst & Young LLP.

Fariborzi, E. y Zahedifard, M. (2012). E-mail Marketing: Advantages, Disadvantages and Improving Techniques. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 2(3), 232-236. Recuperado de <https://www.semanticscholar.org/>

Friedman, M. (1970), *A Theory of the Consumption Function*. India: Mohan Pramlani, Oxford & IBH Publishing Co.

G20. (2018) *Global Partnership for Financial Inclusion (GPMI)*. Recuperado de <https://www.gpmi.org/publications/global-partnership-financial-inclusion-gpmi-argentina-2018-priorities-paper>

Hernández Sampieri, R., Fernández Collado, C., Baptista Lucio, M. (2010). *Metodología de la Investigación*. Distrito Federal, México: McGraw-Hill / Interamericana Editores, S.A. de C.V.

Gujarati, D. y Porter, D. (2010). *Econometría*. Distrito Federal, México: McGraw-Hill / Interamericana Editores, S.A. de C.V.

IIC y Oliver Wyman. (2016). *La revolución Fintech*. Recuperado de <http://www.iic.org/es/media/publication/la-revolucion-fintech>

Instituto Nacional de Estadística y Censos [INDEC]. (2018). *Informes Técnicos: Accesos a Internet* (Vol. 3, nº 44). Recuperado de <https://www.indec.gov.ar/informesdeprensa.asp>

Instituto Nacional de Estadística y Censos [INDEC]. (2019). *Informes Técnicos: Índice de precios al consumidor (IPC)* (Vol. 3, nº 7). Recuperado de https://www.indec.gov.ar/informesdeprensa_antteriores.asp?id_tema_1=3&id_tema_2=5&id_tema_3=31

Jorion, P. (2007). *Financial Risk Manager Handbook*. New Jersey, Estados Unidos: John Wiley & Sons, Inc.

- Kepios. (2019). *Digital 2019: Argentina*. Recuperado de <https://datareportal.com/reports/digital-2019-argentina>
- Keynes, J.M. (1965). *Teoría general de la ocupación, el interés y el dinero*. Distrito Federal, México: Fondo de Cultura Económica.
- Ley 24.144. Boletín Oficial de la República Argentina, Buenos Aires, Argentina, 22 de octubre de 1992.
- Ley 25.326. Boletín Oficial de la República Argentina, Buenos Aires, Argentina, 2 de noviembre de 2000.
- Mercado Libre. (31 de octubre de 2017) *Mercado Crédito: préstamos para potenciar la inclusión financiera*. Recuperado de <https://ideas.mercadolibre.com/ar/noticias/mercado-credito-prestamos-para-potenciar-la-inclusion-financiera/>
- Mester, L. (octubre de 1997). *What's the Point of Credit Scoring?*. Federal Reserve Bank of Philadelphia. Recuperado de <https://pdfs.semanticscholar.org/4ccd/81d64e04ac7cadd9936a703543075fa24846.pdf>
- Morales Castro, J.A y Morales Castro, A (2014). *Crédito y Cobranza*. Distrito Federal, México: Grupo Editorial Patria.
- Morettini, M. (2002). *Principales teorías macroeconómicas sobre el Consumo*. Recuperado de <http://nulan.mdp.edu.ar/1887/1/01486.pdf>
- Oxford latin dictionary (1968) *Creditum*. Oxford, Inglaterra: Oxford University Press.
- Puertas Medina, R. y Martí Selva, M. (2013). *Análisis del credit scoring*. *RAE - Revista de Administração de Empresas*, (Vol 53, Nro 3). Recuperado de <http://www.redalyc.org/articulo.oa?id=155127485011>

Real Academia Española [RAE] (2019) *Buzoneo*. Real Academia Española. Edición del tricentenario. Actualización 2018. Recuperado de <https://dle.rae.es/?id=9NcFAo6>

Real Academia Española [RAE] (2019) *Consumir*. Real Academia Española. Edición del tricentenario. Actualización 2018. Recuperado de <https://dle.rae.es/?id=AT2BY5W>

Samarasinghe, S. (2006). *Neural Networks for Applied Sciences and Engineering*. Nueva York, Estados Unidos: Auerbach Publications.

Schuermann, T. (2004). *What Do We Know About Loss Given Default?*. Wharton Financial Institutions Center, 04(01). Recuperado de <https://ssrn.com/abstract=525702>

Sticco, D. (13 de marzo de 2017). *En la Argentina hay más de tres usuarios de celulares por cada uno con cuenta bancaria*. Recuperado de <https://www.infobae.com/economia/2017/03/13/en-la-argentina-hay-mas-de-3-usuarios-de-celulares-por-cada-uno-con-cuenta-bancaria/>

Twitter. (13 de mayo de 2019). *Política de privacidad de Twitter*. Recuperado de https://cdn.cms-twdigitalassets.com/content/dam/legal-twitter/site-assets/privacy-page-gdpr/PP_EN_Q22018-April_ES.pdf

Anexo I

| Preguntas de investigación. | Objetivos específicos | Fuente de datos/Instrumento de recolección. | Metodología/Técnicas de procesamiento para analizar los datos |
|---|--|---|--|
| ¿Qué variables se consideran relevantes para la evaluación crediticia online en el segmento de consumo? | Indagar variables que se consideren actualmente relevantes para la evaluación crediticia online en el segmento de créditos de consumo. | Fuente primaria: Entrevista a personal de créditos Banco Provincia. Fuente primaria: Entrevista a personal de créditos Banco Columbia. | Cualitativo, descriptivo, transversal, no experimental. Preguntas abiertas |
| ¿Qué variables utilizadas en el mundo digital se pueden incorporar a futuros modelos de scoring crediticio? | Testear variables sociales que puedan ser incorporadas a futuros modelos de scoring crediticio. | Fuente primaria: Base de clientes de Banco Columbia. Fuente primaria: Twitter. | Cuantitativo, hipótesis correlacional, estadísticas, de correlación. |
| ¿Qué herramientas utilizadas en el entorno Fintech pueden aplicarse al mercado crediticio de consumo? | Identificar herramientas utilizadas en el entorno Fintech que sean aplicables al mercado crediticio de consumo. | Fuente primaria: Entrevista a representantes de cámara Fintech. | Cualitativo, descriptivo, transversal, no experimental. Preguntas abiertas |

Anexo II

Formato de tabla de la Central de Deudores del Sistema Financiero – CENDEU
(BCRA)

| N° CAMPO | NOMBRE | TIPO | LONGITUD | OBSERVACIONES |
|-----------------|------------------------|-------------|-----------------|--|
| 1 | Código de entidad | Numérico | 5 | Código de la entidad |
| 2 | Fecha de información | Numérico | 6 | AAAAMM |
| 3 | Tipo de identificación | Numérico | 2 | 11 = clave de identificación fiscal (CUIT, CUIL o CDI) |
| 4 | N° de identificación | Carácter | 11 | |
| 5 | Denominación | Carácter | 55 | |
| 6 | Situación | Numérico | 2 | |
| 7 | Monto | Numérico | 12 | Once enteros y un decimal |

Anexo III

Tabla I – Variables utilizadas

| Variable | Descripción | Tipo de dato |
|------------------------------------|---|--------------|
| Idscoring | Identificación del cliente. Clave primaria. | Número |
| FechaProceso | Fecha en la que se evaluó la solicitud | Fecha |
| SEXO | Codificación del sexo del cliente con M = 0, F = 1 | Binario |
| Edad | Edad del cliente | Entero |
| Actividad | Codificación de actividad. Variable categórica. | Entero |
| Actividad 1 | Variable dummy para empleados privados | Binario |
| Actividad 2 | Variable dummy para autónomos o monotributistas | Binario |
| Actividad 3 | Variable dummy para jubilados | Binario |
| Actividad 4 | Variable dummy para empleado público nacional | Binario |
| Actividad 5 | Variable dummy para empleado público provincial o municipal | Binario |
| Actividad 6 | Variable dummy para servicio doméstico | Binario |
| Actividad 7 | Variable dummy para pensionados | Binario |
| Actividad 8 | Variable dummy para uniformados | Binario |
| Actividad 9 | Variable dummy para contratados privados | Binario |
| Actividad 31 | Variable dummy para cuentapropistas | Binario |
| Etapa laboral | Determina si el cliente es Activo = 1, Pasivo = 0 | Binario |
| Sucursal | Codificación de sucursal | Entero |
| ProvinciaID | Codificación de provincia de residencia | Entero |
| Region | Agrupar por regiones: NOA, NEA, CUYO, SUR, AMBA, CENTRO | Cadena |
| Tipo de operación | Codificación de la operación con 1 = alta y 0 = renovación | Binario |
| IngresoUsado | Ingresos del cliente | Entero |
| Presenta recibo | No = 0 y Si = 1 | Binario |
| Tipo de pago | Pago con Débito = 1 y Pago con cuponera (voluntario) = 0 | Binario |
| Producto sistema financiero | Codificación del producto sistema financiero. | Entero |
| PSF13 | Variable dummy para activos de pago voluntario | Binario |
| PSF23 | Variable dummy para jubilados con débito en cuenta | Binario |
| PSF31 | Variable dummy para activos con débito en cuenta | Binario |
| PSF32 | Variable dummy para nóminas de empleados | Binario |
| PSF36 | Variable dummy para pasivos de pago voluntario | Binario |
| PSF58 | Variable dummy para empleados del Banco | Binario |
| PSF63 | Variable dummy para pasivos con débito en cuenta | Binario |
| Capital solicitado | Monto solicitado en la operación. | Real |
| Ratio capital ingreso | División entre capital solicitado e ingreso usado. | Real |
| Plazo (meses) | Duración del compromiso crediticio | Entero |
| Sit actual BCRA | Calificación crediticia actual del cliente | Entero |
| Exposición | Calificación en el sistema financiero. 1 si está, 0 si no lo está | Binario |

| | | |
|---------------------------------|--|--------|
| Cantidad Bancos actual | Cantidad de entidades con las que opera actualmente | Entero |
| Exposicion actual BCRA | Sumatoria del saldo de deuda actual en el sistema financiero. | Real |
| Ratio exposición ingreso | Cuanto debe en el sistema financiero en relación a lo que dice ganar. Variable construida. | Real |
| % sit 1 actual | Porcentaje de exposición calificada con 1 actualmente. Variable construida | Real |
| %sit 2 actual | Porcentaje de exposición calificada con 2 actualmente. Variable construida | Real |
| Referencias Comerciales | Cantidad de referencias comerciales negativas en el sistema financiero. | Entero |
| Sit BCRA 6m | Peor calificación crediticia del cliente en los últimos 6 meses | Entero |
| Bancos 6m | Máxima cantidad de entidades con las que operó en los últimos 6 meses | Entero |
| Exposicion 6m BCRA | Mayor sumatoria mensual de saldo de deuda en los últimos 6 meses | Entero |
| Cant sit 1 6m | Máxima cantidad mensual de entidades que lo consideran con situación 1 en los últimos 6 meses | Entero |
| Cant sit 2 6m | Máxima cantidad mensual de entidades que lo consideran con situación 2 en los últimos 6 meses | Entero |
| Cant sit >2 6m | Máxima cantidad mensual de entidades que lo consideran con situación mayor a 2 en los últimos 6 meses | Entero |
| Sit BCRA 12m | Peor calificación crediticia del cliente en los últimos 12 meses | Entero |
| Bancos 12m | Máxima cantidad de entidades con las que operó en los últimos 12 meses | Entero |
| Exposicion 12m BCRA | Mayor sumatoria mensual de saldo de deuda en los últimos 12 meses | Entero |
| Cant sit 1 12m | Máxima cantidad mensual de entidades que lo consideran con situación 1 en los últimos 12 meses | Entero |
| Cant sit 2 12m | Máxima cantidad mensual de entidades que lo consideran con situación 2 en los últimos 12 meses | Entero |
| Cant sit >2 12m | Máxima cantidad mensual de entidades que lo consideran con situación mayor a 2 en los últimos 12 meses | Entero |
| Sit BCRA 24m | Máxima cantidad de entidades con las que operó en los últimos 24 meses | Entero |
| Bancos 24m | Mayor sumatoria mensual de saldo de deuda en los últimos 24 meses | Entero |
| Exposicion 24m BCRA | Máxima cantidad mensual de entidades que lo consideran con situación 1 en los últimos 24 meses | Entero |
| Cant sit 1 24m | Máxima cantidad mensual de entidades que lo consideran con situación 2 en los últimos 24 meses | Entero |
| Cant sit 2 24m | Máxima cantidad mensual de entidades que lo consideran con situación mayor a 2 en los últimos 24 meses | Entero |

| | | |
|---------------------------|--|--------|
| Cant sit >2 24m | Máxima cantidad de entidades con las que operó en los últimos 24 meses | Entero |
|---------------------------|--|--------|

Tabla II – Estadísticos descriptivos por variable.

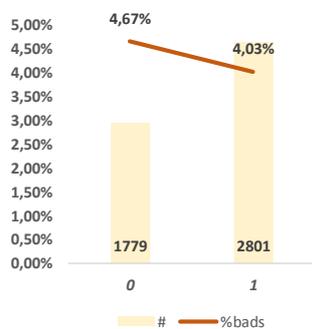
| Variable | Min | Max | Media | Moda | Cantidad | Vacíos |
|------------------------------------|------------|------------|--------------|-------------|-----------------|---------------|
| Sexo | 0 | 1 | 0,61 | 1,00 | 4580 | 0 |
| Edad | 21 | 84 | 57 | 60 | 4580 | 0 |
| Actividad | 1 | 31 | N/A | 3 | 4580 | 0 |
| Actividad 1 | 0 | 1 | 23,65% | 0 | 4580 | 0 |
| Actividad 2 | 0 | 1 | 5,68% | 0 | 4580 | 0 |
| Actividad 3 | 0 | 1 | 51,18% | 1 | 4580 | 0 |
| Actividad 4 | 0 | 1 | 1,24% | 0 | 4580 | 0 |
| Actividad 5 | 0 | 1 | 7,40% | 0 | 4580 | 0 |
| Actividad 6 | 0 | 1 | 1,90% | 0 | 4580 | 0 |
| Actividad 7 | 0 | 1 | 2,49% | 0 | 4580 | 0 |
| Actividad 8 | 0 | 1 | 0,66% | 0 | 4580 | 0 |
| Actividad 9 | 0 | 1 | 0,35% | 0 | 4580 | 0 |
| Actividad 31 | 0 | 1 | 5,46% | 0 | 4580 | 0 |
| Etapa laboral | 0 | 1 | 46,33% | 0 | 4580 | 0 |
| Sucursal | 0 | 1 | 0,61 | 1 | 4580 | 0 |
| ProvinciaID | 0 | 1 | 0,61 | 1 | 4580 | 0 |
| Region | N/A | N/A | N/A | AMBA | 4580 | 0 |
| Tipo de operación | 0 | 1 | 79,45% | 1 | 4580 | 0 |
| IngresoUsado | 2483,07 | 95000 | 14.320,15 | 4810 | 4580 | 0 |
| Presenta recibo | 0 | 1 | 64,91% | 1 | 4580 | 0 |
| Tipo de pago | 0 | 1 | 26,05% | 0 | 4580 | 0 |
| Producto sistema financiero | 13 | 63 | N/A | 13 | 4580 | 0 |
| PSF13 | 0 | 1 | 48,80% | 0 | 4580 | 0 |
| PSF23 | 0 | 1 | 25,70% | 0 | 4580 | 0 |
| PSF31 | 0 | 1 | 2,51% | 0 | 4580 | 0 |
| PSF32 | 0 | 1 | 0,33% | 0 | 4580 | 0 |
| PSF36 | 0 | 1 | 21,94% | 0 | 4580 | 0 |
| PSF58 | 0 | 1 | 0,04% | 0 | 4580 | 0 |
| PSF63 | 0 | 1 | 0,68% | 0 | 4580 | 0 |
| Capital solicitado | 2.500 | 300.000 | 39.292 | 10.000 | 4580 | 0 |
| Ratio capital ingreso | 12% | 3226% | 3,79 | 1 | 4580 | 0 |
| Plazo (meses) | 6 | 60 | 28,15 | 24 | 4580 | 0 |
| Sit actual BCRA | 0 | 2 | N/A | 1 | 4580 | 0 |
| Exposicion | 0 | 1 | 0,87 | 1 | 4580 | 0 |
| Cantidad Bancos actual | 0 | 14 | 2,64 | 2 | 4580 | 0 |
| Exposicion actual BCRA | 0 | 913.258 | 58.155 | 0 | 4580 | 0 |
| Ratio exposicion ingreso | 0 | 8371% | 3,81 | 0 | 4580 | 0 |
| % sit 1 actual | 0 | 100% | 0,87 | 1 | 4580 | 0 |
| %sit 2 actual | 0 | 100% | 0,00 | 0 | 4580 | 0 |

| | | | | | | |
|--------------------------------|---|---------|--------|---|------|---|
| Referencias Comerciales | 0 | 3 | 6% | 0 | 4580 | 0 |
| Sit BCRA 6m | 0 | 5 | 0,92 | 1 | 4580 | 0 |
| Bancos 6m | 0 | 17 | 2,79 | 2 | 4580 | 0 |
| Exposicion 6m BCRA | 0 | 981.866 | 67.346 | 0 | 4580 | 0 |
| Cant sit 1 6m | 0 | 17 | 2,74 | 2 | 4580 | 0 |
| Cant sit 2 6m | 0 | 3 | 0,04 | 0 | 4580 | 0 |
| Cant sit >2 6m | 0 | 2 | 0,01 | 0 | 4580 | 0 |
| Sit BCRA 12m | 0 | 5 | 1,14 | 1 | 4580 | 0 |
| Bancos 12m | 0 | 19 | 3 | 2 | 4580 | 0 |
| Exposicion 12m BCRA | 0 | 913.258 | 61.572 | 0 | 4580 | 0 |
| Cant sit 1 12m | 0 | 19 | 2,97 | 2 | 4580 | 0 |
| Cant sit 2 12m | 0 | 4 | 0,09 | 0 | 4580 | 0 |
| Cant sit >2 12m | 0 | 6 | 0,08 | 0 | 4580 | 0 |
| Sit BCRA 24m | 0 | 6 | 1,40 | 1 | 4580 | 0 |
| Bancos 24m | 0 | 23 | 3,58 | 2 | 4580 | 0 |
| Exposicion 24m BCRA | 0 | 913.258 | 63.630 | 0 | 4580 | 0 |
| Cant sit 1 24m | 0 | 21 | 3,21 | 2 | 4580 | 0 |
| Cant sit 2 24m | 0 | 4 | 0,16 | 0 | 4580 | 0 |
| Cant sit >2 24m | 0 | 8 | 0,21 | 0 | 4580 | 0 |

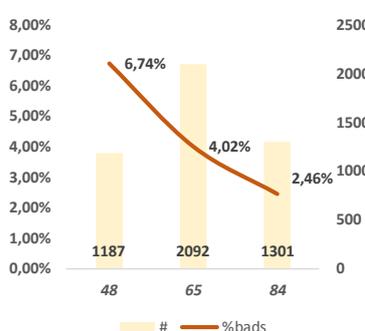
Análisis bivariado de las variables consideradas para realizar modelo tradicional.

Tabla III - Variables sociodemográficas

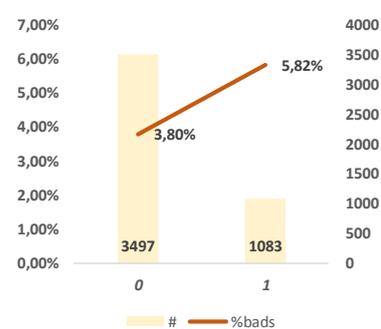
| Sexo | | | Edad | | | Actividad 1 | | |
|--------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|
| Valor | # | %bads | Valor | # | %bads | Valor | # | %bads |
| 0 | 1779 | 4,67% | 48 | 1187 | 6,74% | 0 | 3497 | 3,80% |
| 1 | 2801 | 4,03% | 65 | 2092 | 4,02% | 1 | 1083 | 5,82% |
| Total | 4580 | 4,28% | Total | 4580 | 4,28% | Total | 4580 | 4,28% |



La variable tiene materialidad
Hay mas mujeres que varones
Las mujeres tienen mejor perfil



La variable tiene materialidad
La edad mejora el comportamiento



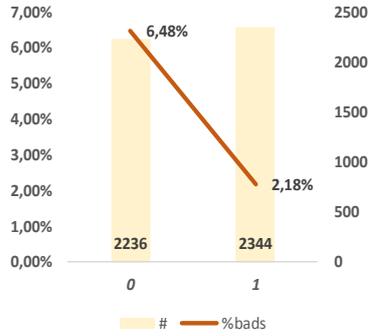
La variable tiene materialidad
Tener act 1 es peor que no tenerla

| Actividad 2 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4320 | 4,03% |
| 1 | 260 | 8,46% |
| Total | 4580 | 4,28% |

| Actividad 3 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 2236 | 6,48% |
| 1 | 2344 | 2,18% |
| Total | 4580 | 4,28% |

| Actividad 4 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4523 | 4,31% |
| 1 | 57 | 1,75% |
| Total | 4580 | 4,28% |

La variable no tiene materialidad



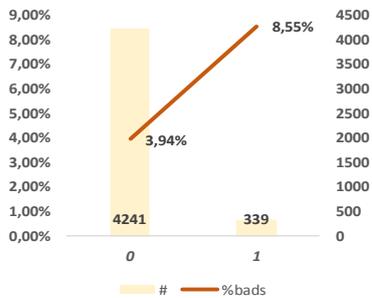
La variable no tiene materialidad

La variable tiene materialidad
Tener act 3 es mejor que no tenerla

| Actividad 5 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4241 | 3,94% |
| 1 | 339 | 8,55% |
| Total | 4580 | 4,28% |

| Actividad 6 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4493 | 4,25% |
| 1 | 87 | 5,75% |
| Total | 4580 | 4,28% |

| Actividad 7 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4466 | 4,25% |
| 1 | 114 | 5,26% |
| Total | 4580 | 4,28% |



La variable no tiene materialidad

La variable no tiene materialidad

La variable tiene materialidad
Tener act 5 es peor que no tenerla

| Actividad 8 | | | Actividad 9 | | | Actividad 31 | | |
|--------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|
| Valor | # | %bads | Valor | # | %bads | Valor | # | %bads |
| 0 | 4550 | 4,26% | 0 | 4564 | 4,27% | 0 | 4330 | 4,16% |
| 1 | 30 | 6,67% | 1 | 16 | 6,25% | 1 | 250 | 6,40% |
| Total | 4580 | 4,28% | Total | 4580 | 4,28% | Total | 4580 | 4,28% |

La variable no tiene materialidad

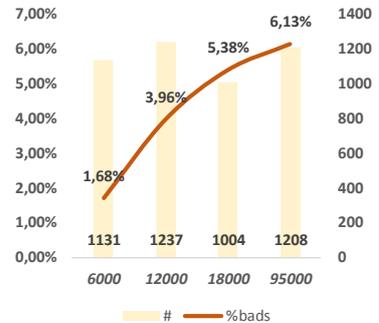
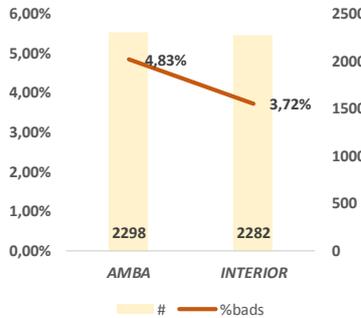
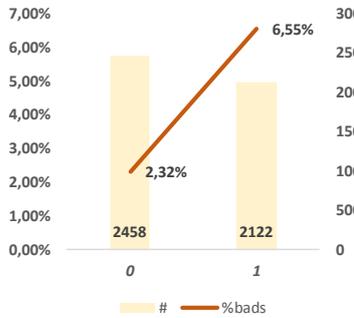
La variable no tiene materialidad

La variable no tiene materialidad

| Etapa laboral | | |
|---------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 2458 | 2,32% |
| 1 | 2122 | 6,55% |
| Total | 4580 | 4,28% |

| Region | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| AMBA | 2298 | 4,83% |
| INTERIOR | 2282 | 3,72% |
| Total | 4580 | 4,28% |

| IngresoUsado | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 6000 | 1131 | 1,68% |
| 12000 | 1237 | 3,96% |
| 18000 | 1004 | 5,38% |
| 95000 | 1208 | 6,13% |
| Total | 4580 | 4,28% |

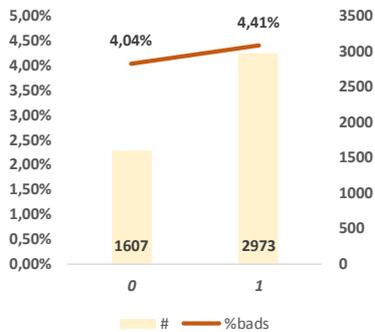


La variable tiene materialidad
Los pasivos pagan mejor que los activos

La variable tiene materialidad
Interior del pais paga mejor que AMBA

La variable tiene materialidad
A mayor ingreso, peor comportamiento

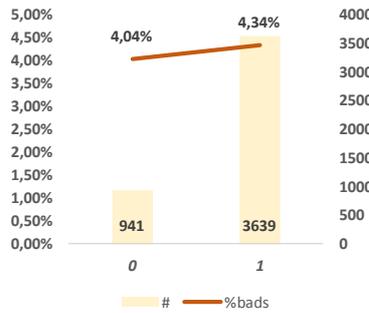
| Presenta recibo | | |
|-----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 1607 | 4,04% |
| 1 | 2973 | 4,41% |
| Total | 4580 | 4,28% |



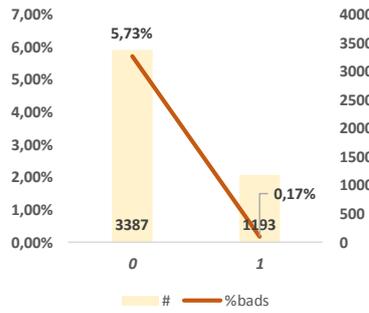
La variable tiene materialidad
Mucha mas gente presenta recibo
Los que lo hacen, son peores pagadores

Tabla IV - Variables de producto

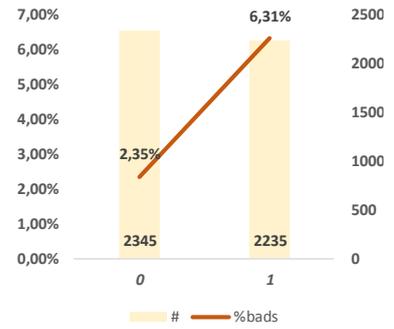
| Tipo de operación | | | Tipo de pago | | | PSF13 | | |
|-------------------|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|
| Valor | # | %bads | Valor | # | %bads | Valor | # | %bads |
| 0 | 941 | 4,04% | 0 | 3387 | 5,73% | 0 | 2345 | 2,35% |
| 1 | 3639 | 4,34% | 1 | 1193 | 0,17% | 1 | 2235 | 6,31% |
| Total | 4580 | 4,28% | Total | 4580 | 4,28% | Total | 4580 | 4,28% |



La variable tiene materialidad
Hay muchas mas altas que renovaciones en l
No se puede predecir comportamiento



La variable tiene materialidad
La mayoría elige pago voluntario
Los que lo hacen, son peores pagadores

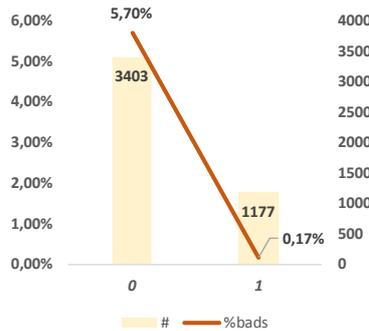


La variable tiene materialidad
La mitad de la muestra pertenece al grupo
Quienes lo hacen, son peores pagadores

| PSF23 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 3403 | 5,70% |
| 1 | 1177 | 0,17% |
| Total | 4580 | 4,28% |

| PSF31 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4465 | 4,17% |
| 1 | 115 | 8,70% |
| Total | 4580 | 4,28% |

| PSF32 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4565 | 4,29% |
| 1 | 15 | 0,00% |
| Total | 4580 | 4,28% |



La variable no tiene materialidad

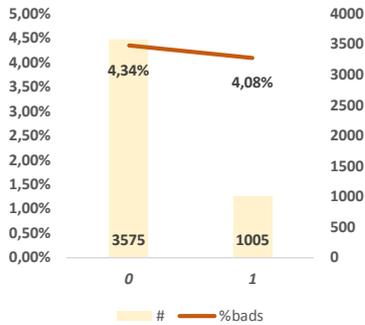
La variable no tiene materialidad

La variable tiene materialidad
Se comporta de manera similar al PCD
Tiene correlacion directa, por el producto

| PSF36 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 3575 | 4,34% |
| 1 | 1005 | 4,08% |
| Total | 4580 | 4,28% |

| PSF58 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4578 | 4,28% |
| 1 | 2 | 0,00% |
| Total | 4580 | 4,28% |

| PSF63 | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4549 | 4,26% |
| 1 | 31 | 6,45% |
| Total | 4580 | 4,28% |



La variable no tiene materialidad

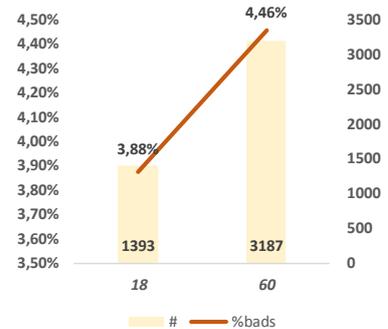
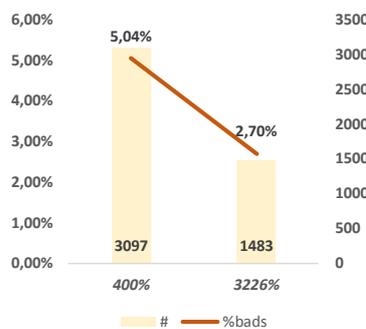
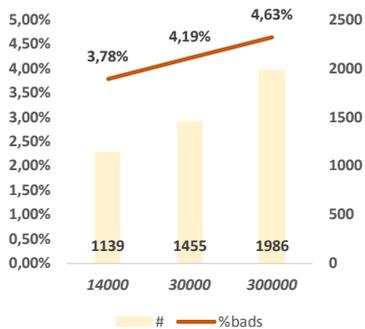
La variable no tiene materialidad

La variable tiene materialidad
Representa al 25% de la muestra
No se puede predecir comportamiento

| Capital solicitado | | |
|--------------------|-------------|--------------|
| Valor | # | %bads |
| 14000 | 1139 | 3,78% |
| 30000 | 1455 | 4,19% |
| 300000 | 1986 | 4,63% |
| Total | 4580 | 4,28% |

| Ratio capital ingreso | | |
|-----------------------|-------------|--------------|
| Valor | # | %bads |
| 400% | 3097 | 5,04% |
| 3226% | 1483 | 2,70% |
| Total | 4580 | 4,28% |

| Plazo (meses) | | |
|---------------|-------------|--------------|
| Valor | # | %bads |
| 18 | 1393 | 3,88% |
| 60 | 3187 | 4,46% |
| Total | 4580 | 4,28% |



La variable tiene materialidad
Se observa tendencia creciente al no pago

La variable tiene materialidad
Mas de cuatro ingresos, peor pagador

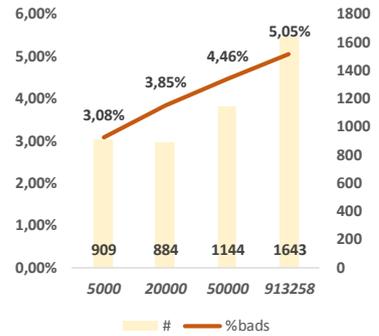
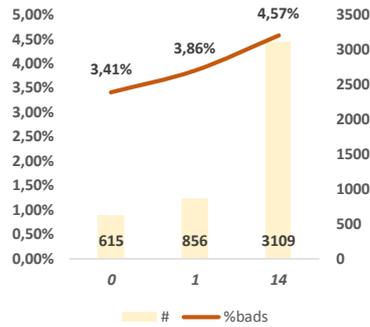
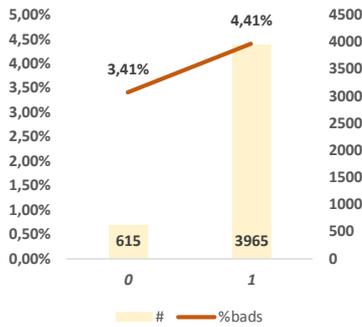
La variable tiene materialidad
Los que sacan a corto plazo son mejores

Tabla V - Variables de comportamiento crediticio

| Exposicion | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 615 | 3,41% |
| 1 | 3965 | 4,41% |
| Total | 4580 | 4,28% |

| Cantidad Bancos actual | | |
|------------------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 615 | 3,41% |
| 1 | 856 | 3,86% |
| 14 | 3109 | 4,57% |
| Total | 4580 | 4,28% |

| Exposicion actual BCRA | | |
|------------------------|-------------|--------------|
| Valor | # | %bads |
| 5000 | 909 | 3,08% |
| 20000 | 884 | 3,85% |
| 50000 | 1144 | 4,46% |
| 913258 | 1643 | 5,05% |
| Total | 4580 | 4,28% |



La variable tiene materialidad
Quienes tienen deuda bancaria son peores

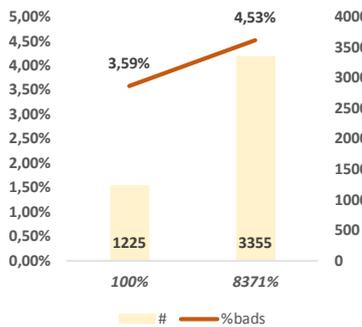
La variable tiene materialidad
A mayor cantidad de bancos, peor pagador
Esto se exceptua para alta cantidad

La variable tiene materialidad
A mayor exposición en dinero, peor pagador

| Ratio exposicion ingreso | | |
|--------------------------|-------------|--------------|
| Valor | # | %bads |
| 100% | 1225 | 3,59% |
| 8371% | 3355 | 4,53% |
| Total | 4580 | 4,28% |

| % sit 1 actual | | |
|----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 616 | 3,41% |
| 0,99 | 10 | 0,00% |
| 1 | 3954 | 4,43% |
| Total | 4580 | 4,28% |

| %sit 2 actual | | |
|---------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4569 | 4,29% |
| 0 | 0 | 0,00% |
| 0 | 0 | 0,00% |
| 1 | 11 | 0,00% |
| Total | 4580 | 4,28% |



Se solapa con "exposición"

La variable no tiene materialidad

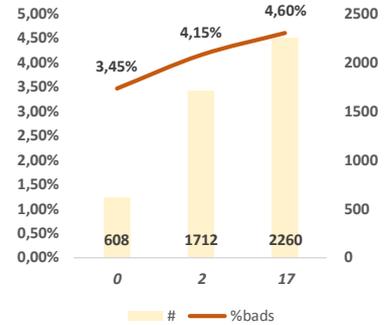
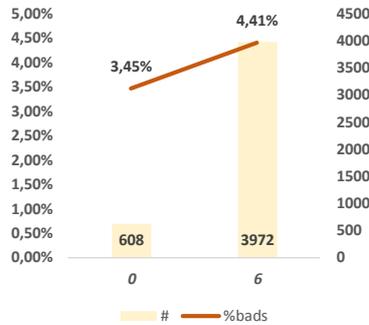
La variable tiene materialidad
Si la exposicion es menor al 100%, mejora

| Referencias Comerciales | | |
|-------------------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4331 | 4,48% |
| 1 | 214 | 0,93% |
| 2 | 31 | 0,00% |
| 3 | 4 | 0,00% |
| Total | 4580 | 4,28% |

| Sit BCRA 6m | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 608 | 3,45% |
| 6 | 3972 | 4,41% |
| Total | 4580 | 4,28% |

| Bancos 6m | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 608 | 3,45% |
| 2 | 1712 | 4,15% |
| 17 | 2260 | 4,60% |
| Total | 4580 | 4,28% |

La variable no tiene materialidad



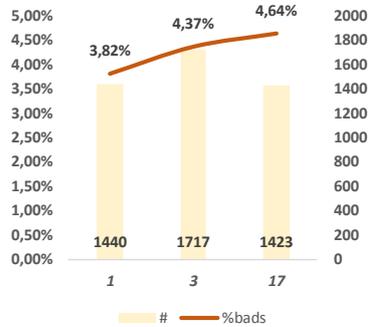
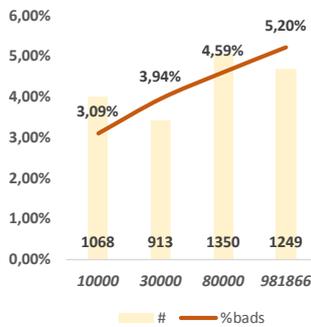
La variable tiene materialidad
Se podrían unificar las sits recientes

| Exposicion 6m BCRA | | |
|--------------------|-------------|--------------|
| Valor | # | %bads |
| 10000 | 1068 | 3,09% |
| 30000 | 913 | 3,94% |
| 80000 | 1350 | 4,59% |
| 981866 | 1249 | 5,20% |
| Total | 4580 | 4,28% |

| Cant sit 1 6m | | |
|---------------|-------------|--------------|
| Valor | # | %bads |
| 1 | 1440 | 3,82% |
| 3 | 1717 | 4,37% |
| 17 | 1423 | 4,64% |
| Total | 4580 | 4,28% |

La variable tiene materialidad
Se comporta igual que cant bancos actual

| Cant sit 2 6m | | |
|---------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4417 | 4,23% |
| 1 | 152 | 5,26% |
| 3 | 11 | 9,09% |
| Total | 4580 | 4,28% |



La variable no tiene materialidad

La variable tiene materialidad
A mayor exposición, peor pagador

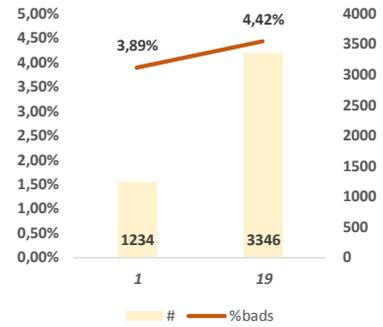
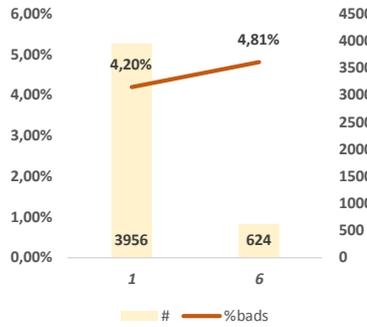
La variable tiene materialidad
A mayor cantidad de bancos, peor pagador

| Cant sit >2 6m | | |
|----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4537 | 4,32% |
| 1 | 42 | 0,00% |
| 2 | 1 | 0,00% |
| Total | 4580 | 4,28% |

| Sit BCRA 12m | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 1 | 3956 | 4,20% |
| 6 | 624 | 4,81% |
| Total | 4580 | 4,28% |

| Bancos 12m | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 1 | 1234 | 3,89% |
| 19 | 3346 | 4,42% |
| Total | 4580 | 4,28% |

La variable no tiene materialidad



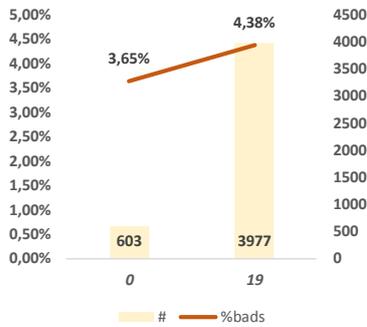
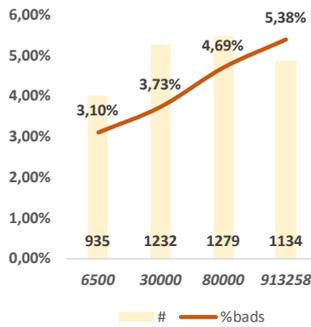
No se si tiene materialidad
A peor situacion BCRA, peor pagador

La variable tiene materialidad
A mayor cantidad de bancos, peor pagador
Poca pendiente

| Exposicion 12m BCRA | | |
|---------------------|-------------|--------------|
| Valor | # | %bads |
| 6500 | 935 | 3,10% |
| 30000 | 1232 | 3,73% |
| 80000 | 1279 | 4,69% |
| 913258 | 1134 | 5,38% |
| Total | 4580 | 4,28% |

| Cant sit 1 12m | | |
|----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 603 | 3,65% |
| 19 | 3977 | 4,38% |
| Total | 4580 | 4,28% |

| Cant sit 2 12m | | |
|----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4207 | 4,11% |
| 1 | 326 | 4,91% |
| 2 | 35 | 14,29% |
| 4 | 12 | 16,67% |
| Total | 4580 | 4,28% |



La variable no tiene materialidad

La variable tiene materialidad
A mayor exposición, peor pagador

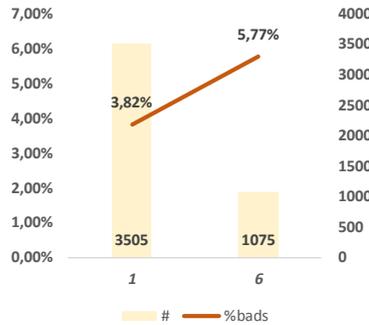
La variable tiene materialidad
No marca tendencia

| Cant sit >2 12m | | |
|-----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4279 | 4,39% |
| 1 | 244 | 2,05% |
| 2 | 35 | 5,71% |
| 6 | 22 | 4,55% |
| Total | 4580 | 4,28% |

| Sit BCRA 24m | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 1 | 3505 | 3,82% |
| 6 | 1075 | 5,77% |
| Total | 4580 | 4,28% |

| Bancos 24m | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 474 | 2,95% |
| 23 | 4106 | 4,43% |
| Total | 4580 | 4,28% |

La variable no tiene materialidad



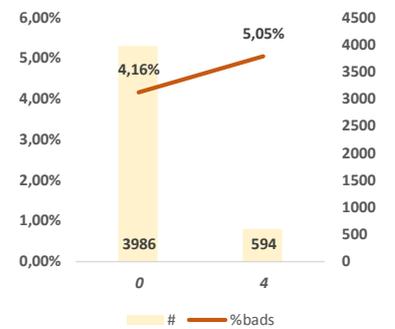
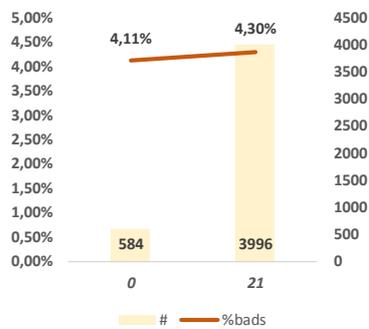
La variable tiene materialidad
Si tuvo atrasos en 2 años, es peor pagador

| Exposicion 24m BCRA | | |
|---------------------|-------------|--------------|
| Valor | # | %bads |
| 10000 | 1088 | 3,22% |
| 30000 | 998 | 3,81% |
| 80000 | 1323 | 4,61% |
| 913258 | 1171 | 5,29% |
| Total | 4580 | 4,28% |

| Cant sit 1 24m | | |
|----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 584 | 4,11% |
| 21 | 3996 | 4,30% |
| Total | 4580 | 4,28% |

La variable tiene materialidad
A mayor exposicion, peor pagador

| Cant sit 2 24m | | |
|----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 3986 | 4,16% |
| 4 | 594 | 5,05% |
| Total | 4580 | 4,28% |

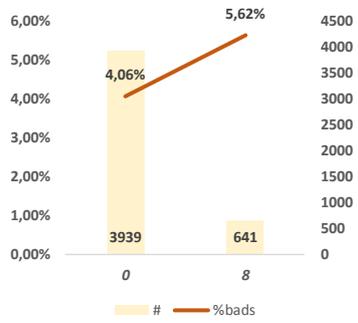


La variable tiene materialidad
A mayor exposicion, peor pagador

La variable tiene materialidad
No muestra tendencia

La variable tiene materialidad
Quienes tienen exposicion pagan peor

| Cant sit >2 24m | | |
|-----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 3939 | 4,06% |
| 8 | 641 | 5,62% |
| Total | 4580 | 4,28% |



La variable tiene materialidad
A mayor cantidad de sit >2, peor pagador

Anexo IV

Tabla I – Variables utilizadas

| Variable | Descripción | Tipo de dato |
|-------------------------------|--|--------------|
| nombre en mail | Variable dicotómica que verifica que coincidan las cadenas correspondientes al “nombre” de la solicitud con el propietario del mail, con 0 para la negativa y 1 para el caso afirmativo. | Binario |
| Apellido en mail | Variable dicotómica que verifica que coincidan las cadenas correspondientes al “apellido” de la solicitud con el propietario del mail, con 0 para la negativa y 1 para el caso afirmativo. | Binario |
| Repetición de mail | Sumatoria del mail declarado por el cliente en la base. A mayor cantidad de personas utilizando el mismo mail, mayor probabilidad de que no corresponda a su e-mail real. | Entero |
| Email repetido | Verificación de que algún otro cliente haya usado el mismo mail. 0 para negativo y 1 para afirmativo. | Binario |
| Es mail basura? | Verificación manual de que el string del cuerpo del mail no asuma un comportamiento irracional. Por Ej: “aaaa@abc.com” 0 para negativo y 1 para afirmativo. | Binario |
| Es gmail | Verificación del servidor de correo electrónico @gmail. 0 para negativo y 1 para afirmativo. | Binario |
| Es hotmail | Verificación del servidor de correo electrónico @hotmail. 0 para negativo y 1 para afirmativo. | Binario |
| Es yahoo | Verificación del servidor de correo electrónico @yahoo. 0 para negativo y 1 para afirmativo. | Binario |
| Es Outlook | Verificación del servidor de correo electrónico @outlook. 0 para negativo y 1 para afirmativo. | Binario |
| Origen servidor | Verificación del origen del servidor de correo electrónico, con 0 para.ar y 1 para extranjeros | Binario |
| Guiones en mail | Variable dicotómica que verifica que existan guiones en las cadenas correspondientes al cuerpo del mail, con 0 para la negativa y 1 para el caso afirmativo. | Binario |
| Punto en mail | Variable dicotómica que verifica que existan puntos en las cadenas correspondientes al cuerpo del mail, con 0 para la negativa y 1 para el caso afirmativo. | Binario |
| Tiene twitter asociado | Variable dicotómica que verifica si el mail declarado por el cliente tiene cuenta de twitter asociada, con 0 para la negativa y 1 para el caso afirmativo. | Binario |
| Coincide cel | Variable dicotómica que verifica si el celular declarado por el bureau de crédito para cada cliente coincide en sus últimas dos cifras con el declarado en su cuenta de twitter asociada. 0 para negativo y 1 para afirmativo. | Binario |
| DeclaraCel | Variable dicotómica que verifica si el cliente declara un teléfono con prefijo 15, de tipo celular, con 0 para la negativa y 1 para el caso afirmativo. | Binario |
| CoincideTelDec | Variable dicotómica que verifica si el teléfono declarado por el cliente coincide con alguno de los detectados por el bureau de crédito. 0 para negativo y 1 para afirmativo. | Binario |

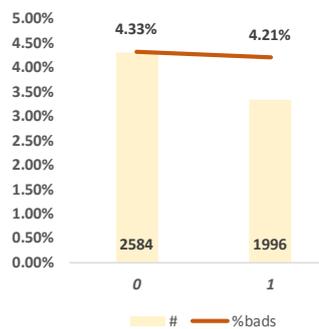
Tabla II – Estadísticos descriptivos por variable.

| Variable | Min | Max | Media | Moda | Cantidad | Vacíos |
|------------------------|-----|-----|--------|------|----------|--------|
| nombre en mail | 0 | 1 | 43.58% | 0 | 4580 | 0 |
| Apellido en mail | 0 | 1 | 48.10% | 0 | 4580 | 0 |
| Email repetido | 1 | 18 | 1.04 | 1 | 4580 | 0 |
| Repeticion de mail | 0 | 1 | 2.29% | 0 | 4580 | 0 |
| Es mail basura? | 0 | 1 | 1.38% | 0 | 4580 | 0 |
| Es gmail | 0 | 1 | 59.56% | 1 | 4580 | 0 |
| Es hotmail | 0 | 1 | 31.68% | 0 | 4580 | 0 |
| Es yahoo | 0 | 1 | 3.97% | 0 | 4580 | 0 |
| Es Outlook | 0 | 1 | 1.59% | 0 | 4580 | 0 |
| Origen servidor | 0 | 1 | 91.46% | 1 | 4580 | 0 |
| Guiones en mail | 0 | 1 | 10.94% | 0 | 4580 | 0 |
| Punto en mail | 0 | 1 | 8.49% | 0 | 4580 | 0 |
| Tiene twitter asociado | 0 | 1 | 24.19% | 0 | 4580 | 0 |
| Coincide cel | 0 | 1 | 1.00% | 0 | 4580 | 0 |
| DeclaraCel | 0 | 1 | 15.02% | 0 | 4580 | 0 |
| CoincideTelDec | 0 | 1 | 60.94% | 1 | 4580 | 0 |

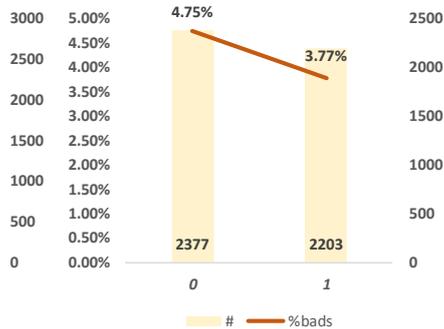
Análisis bivariado de las variables consideradas para realizar modelo digital.

Tabla III - Variables utilizadas

| nombre en mail | | | Apellido en mail | | | Repeticion de mail | | |
|----------------|-------------|--------------|------------------|-------------|--------------|--------------------|-------------|--------------|
| Valor | # | %bads | Valor | # | %bads | Valor | # | %bads |
| 0 | 2584 | 4.33% | 0 | 2377 | 4.75% | 1 | 4475 | 4.27% |
| 1 | 1996 | 4.21% | 1 | 2203 | 3.77% | 2 | 88 | 5.68% |
| Total | 4580 | 4.28% | Total | 4580 | 4.28% | 3 | 6 | 0.00% |
| | | | | | | 18 | 11 | 0.00% |
| | | | | | | Total | 4580 | 4.28% |



La variable tiene materialidad
No muestra tendencia



La variable tiene materialidad
Apellido en el mail mejora comportamiento

La variable no tiene materialidad

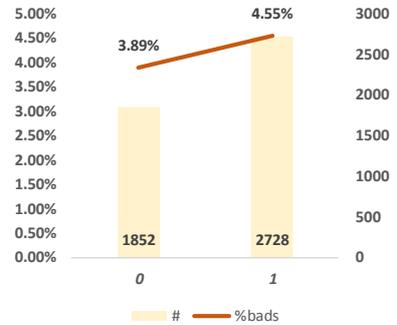
| Email repetido | | |
|----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4475 | 4.27% |
| 1 | 105 | 4.76% |
| Total | 4580 | 4.28% |

| Es mail basura? | | |
|-----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4517 | 4.27% |
| 1 | 63 | 4.76% |
| Total | 4580 | 4.28% |

| Es gmail | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 1852 | 3.89% |
| 1 | 2728 | 4.55% |
| Total | 4580 | 4.28% |

La variable no tiene materialidad

La variable no tiene materialidad

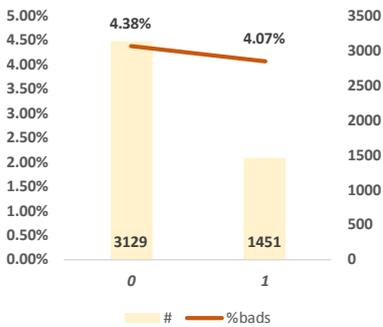


La variable tiene materialidad
Dar casilla gmail te hace peor pagador

| Es hotmail | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 3129 | 4.38% |
| 1 | 1451 | 4.07% |
| Total | 4580 | 4.28% |

| Es yahoo | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4398 | 4.30% |
| 1 | 182 | 3.85% |
| Total | 4580 | 4.28% |

| Es outlook | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4507 | 4.30% |
| 1 | 73 | 2.74% |
| Total | 4580 | 4.28% |



La variable no tiene materialidad

La variable no tiene materialidad

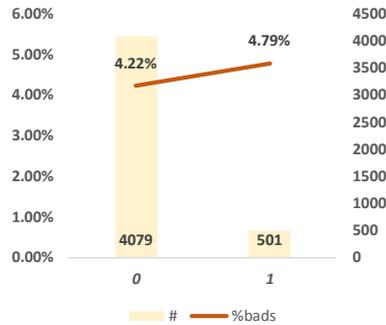
La variable tiene materialidad
No hay tendencia definida

| Origen servidor | | |
|-----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 391 | 4.35% |
| 1 | 4189 | 4.27% |
| Total | 4580 | 4.28% |

| Guiones en mail | | |
|-----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4079 | 4.22% |
| 1 | 501 | 4.79% |
| Total | 4580 | 4.28% |

| Punto en mail | | |
|---------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4191 | 4.25% |
| 1 | 389 | 4.63% |
| Total | 4580 | 4.28% |

La variable no tiene materialidad



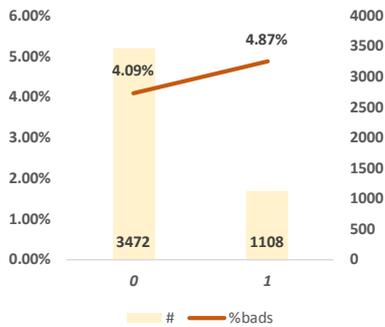
La variable no tiene materialidad

La variable tiene materialidad
Los que usan guion son peores pagadores

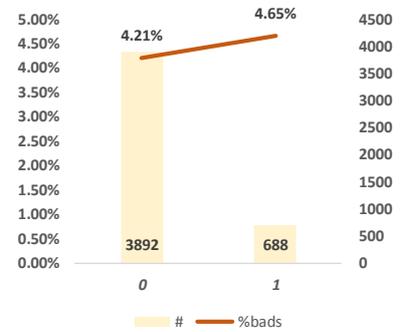
| Tiene twitter asociado | | |
|------------------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 3472 | 4.09% |
| 1 | 1108 | 4.87% |
| Total | 4580 | 4.28% |

| Coincide cel | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 4534 | 4.21% |
| 1 | 46 | 10.87% |
| Total | 4580 | 4.28% |

| Declara Cel | | |
|--------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 3892 | 4.21% |
| 1 | 688 | 4.65% |
| Total | 4580 | 4.28% |



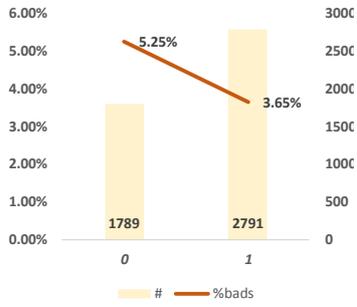
La variable no tiene materialidad



La variable tiene materialidad
Los que asocian twitter son peor pagadores

La variable tiene materialidad
Los que declaran celular pagan peor

| CoincideTelDec | | |
|----------------|-------------|--------------|
| Valor | # | %bads |
| 0 | 1789 | 5.25% |
| 1 | 2791 | 3.65% |
| Total | 4580 | 4.28% |



La variable tiene materialidad
Si coinciden con el bureau pagan mejor