

EDUCACIÓN FINANCIERA Y RIESGO DE CRÉDITO EN MÉXICO*

Eduardo Bello Toledo y Daniel A. Miranda López

Resumen. En los últimos años, en México, se han observado avances encomiables en la bancarización de la población, a pesar de ello el acceso al financiamiento continúa mostrando rezagos. Sin embargo, la constante expansión de la literatura sobre el estudio de las causas y retos de la inclusión financiera han abierto un nuevo panorama hacia variables que podrían incidir en el desempeño financiero de la población, entre las que se encuentra la alfabetización financiera, la cual, podría incrementar los beneficios potenciales que generan mayores niveles de inclusión financiera. En este trabajo se analiza si el nivel de alfabetización financiera conformada por los conocimientos, comportamientos y actitudes financieras incide en la probabilidad de impago de sus obligaciones con instituciones financieras, y si este mismo nivel de alfabetización podría ser útil para realizar un análisis crediticio. Para dar respuesta a estos objetivos se utiliza la información de los últimos dos levantamientos de la Encuesta Nacional de Inclusión Financiera (ENIF 2018 y 2021). Los resultados sugieren que existe una relación negativa entre los niveles de alfabetización financiera y la probabilidad de impago, lo cual es consistente con la literatura, sin embargo, a diferencia de los trabajos previos, los resultados permiten desagregar más el análisis y apuntan a que los beneficios se deben a los buenos comportamientos financieros, por lo que, el fomentar el desarrollo de buenos comportamientos financieros, como el establecimiento de metas a largo plazo y promover el hábito del ahorro, podrían ser fundamentales para formar acreditados que muestren un mejor desempeño en el mercado, lo cual se reflejará en menores costos de financiamiento. El análisis muestra que los comportamientos financieros tienen un poder explicativo similar al que muestran las variables de historial financiero/crediticio, lo cual cobra mayor relevancia cuando se trata de personas no incluidas en el sistema financiero, ya que dicha información podría ser la clave para que este sector de la población encuentre un canal de acceso al financiamiento.

Palabras clave: Alfabetización financiera, financiamiento, riesgo de crédito, probabilidad de impago.

JEL: C10, C52, I22

**/ Esta investigación corresponde a un documento de trabajo. Las opiniones vertidas en este trabajo corresponden únicamente a los autores y no necesariamente reflejan la postura institucional de la CNBV.*

FINANCIAL EDUCATION AND CREDIT RISK IN MEXICO*

Eduardo Bello Toledo y Daniel A. Miranda López

Abstract. In recent years, Mexico has made commendable progress in the banking penetration of the population, although access to financing continues to lag behind. However, the constant expansion of the literature on the study of the causes and challenges of financial inclusion has opened a new panorama towards variables that could affect the financial performance of the population, among which is financial literacy, which, could increase the potential benefits generated by higher levels of financial inclusion. This paper analyzes whether the level of financial literacy, made up of financial knowledge, behaviors and attitudes, affects the probability of non-payment of its obligations with financial institutions, and whether this same level of literacy could be useful for carrying out a credit analysis. To respond to these objectives, information from the last two surveys of the National Survey of Financial Inclusion (ENIF 2018 and 2021) is used. The results suggest that there is a negative relationship between the levels of financial literacy and the probability of default, which is consistent with the literature; however, unlike previous studies, the results allow further disaggregation of the analysis and suggest that the benefits are due to good financial behaviors, therefore, encouraging the development of good financial behaviors, such as establishing long-term goals and promoting the habit of saving, could be essential to train borrowers who show better performance in the market, which will be reflected in lower financing costs. The analysis shows that financial behaviors have an explanatory power similar to that shown by the financial/credit history variables, which becomes more relevant when it comes to people not included in the financial system, since said information could be the key to this sector of the population finds a channel of access to financing.

Keywords: Financial literacy, financing, credit risk, probability of default.

JEL: C10, C52, I22

**/ This document represents a working paper. The opinions expressed in it are responsibility of the authors and do not necessarily reflect the institutional position of the CNBV.*

EDUCACIÓN FINANCIERA Y RIESGO DE CRÉDITO EN MÉXICO*

Eduardo Bello Toledo y Daniel Alejandro Miranda López^o

I. Introducción

La inclusión financiera de la población, definida como el nivel de acceso y uso de los productos y servicios financieros que funcionan bajo una regulación que procure esquemas de protección al consumidor y que promueva la educación financiera para mejorar las capacidades económico-financieras de los usuarios (Comisión Nacional Bancaria y de Valores, 2022), ha sido materia de cada vez más trabajos de investigación (Comisión Nacional Bancaria y de Valores, 2020) debido a que ha mostrado que puede incidir de forma positiva en el bienestar de los hogares, permitiéndoles acceder a los productos y servicios financieros que requieren para satisfacer sus necesidades de consumo y alcanzar sus metas económicas y financieras. Además, a nivel macroeconómico, la literatura ha evidenciado la relación positiva existente entre mayores niveles de inclusión financiera y el crecimiento y desarrollo económico de un país (Barajas et al., 2020; Beck et al., 2009; Cull et al., 2014; Demircuc-Kunt et al., 2017). De esta forma, promover la inclusión financiera es un mecanismo indispensable para contrarrestar los problemas de desigualdad y pobreza, que se han acentuado a pesar del desarrollo del sistema financiero global (Loukoianova & Yang, 2018; Sahay & Chiak, 2020).

En México, se han dado avances encomiables en materia de inclusión financiera. En 2021, según datos de la Encuesta Nacional de Inclusión Financiera (ENIF) 2021, 8 de cada 10 personas adultas habían tenido al menos un producto financiero y 65% habían tenido al menos una cuenta de ahorro en una entidad financiera, porcentaje que se incrementa a 84% en el caso de las empresas de más de 6 empleados, según datos de la Encuesta Nacional de Financiamiento de las Empresas (ENAFIN) 2021. Además de esto, cada vez es más común el uso de nuevas tecnologías entre quienes participan en el sistema financiero. En el mismo período, 1 de cada 3 empresas contaba con banca móvil y 6 de cada 10 utilizaba la banca por internet para realizar sus

**/ Esta investigación corresponde a un documento de trabajo. Las opiniones vertidas en este trabajo corresponden únicamente a los autores y no necesariamente reflejan la postura institucional de la CNBV.*

^o Agradecemos el inestimable apoyo brindado por la Dra. Lucía Buenrostro, su guía, cuestionamientos y comentarios fueron esenciales para la conclusión exitosa de este documento. Asimismo, agradecemos las revisiones y comentarios realizados por Cristóbal Domínguez, Sofía Huidobro y el equipo de la Dirección General para el Acceso a Servicios Financieros, los cuales ayudaron a corregir errores, esclarecer algunos resultados y ampliar la discusión para robustecer la investigación.



operaciones financieras. Mientras que, entre los adultos con cuenta de ahorro, más de la mitad ya realiza sus operaciones a través del celular, proporción que se incrementa significativamente en la población más joven, lo cual sugiere que la digitalización de los servicios financieros será determinante para que las nuevas generaciones se incorporen al sistema financiero.

A pesar de estos avances, el acceso al financiamiento continúa mostrando rezago tanto entre la población como en el sector empresarial, lo cual, podría reducir los beneficios potenciales que podrían acarrear mayores niveles de inclusión financiera. En particular, el crédito permite que tanto personas como empresas puedan obtener los recursos necesarios para realizar sus actividades, expandir sus negocios, satisfacer su consumo, planificar su operación y, en general, alcanzar las metas que la limitación natural de los recursos financieros podría obstaculizar (Banerjee et al., 2015). A pesar de los beneficios potenciales que podría acarrear una amplia participación en el mercado crediticio, en México la penetración del financiamiento es acotada. Según la ENIF 2021, más de la mitad de la población adulta no contaba con financiamiento y solo 33% contaba con crédito de una institución financiera regulada. Además, 53% de la población adulta nunca había contado con algún crédito de una institución financiera, mientras que, entre las empresas 53% nunca había tenido o solicitado algún financiamiento (ENAFIN, 2021), y dichos niveles han cambiado poco en los últimos años, a pesar de que el sistema financiero ha continuado expandiéndose, por la llegada de nuevos participantes y por la mayor adopción de nuevas tecnologías para la provisión de servicios y el diseño de productos financieros.

Lo anterior, es evidencia de que el incremento de la oferta de los servicios financieros no se traduce necesariamente en mayor inclusión financiera de forma directa, ya que existen diversos factores que podrían limitar la participación de un mayor número de personas y empresas en el sistema financiero, limitándolos de los beneficios que este provee, restringiendo su acceso al financiamiento y, potencialmente, afectando el crecimiento y desarrollo económico en el país. Sin duda, uno de los factores que ha sido más ampliamente estudiado en la literatura económica es el riesgo de crédito. Cada vez que una institución financiera analiza una solicitud de crédito se enfrenta a la posibilidad de que el solicitante no cumpla con la obligación de pagar si se decide otorgarle financiamiento, y esto afecte el balance financiero de la institución (Chorafas, 2000).

La existencia de asimetrías de información entre los solicitantes de financiamiento y las instituciones de crédito ha propiciado que las instituciones financieras busquen establecer métricas y metodologías que contribuyan a una adecuada medición y gestión de estos riesgos, ya que, una

toma excesiva de riesgo podría poner en peligro su solvencia y resiliencia, y afectar la confianza que los usuarios tienen en los intermediarios financieros, la cual, es fundamental para que el sistema financiero funcione de forma eficiente. Por otra parte, mitigar esta asimetría mediante métodos estadísticos ha ocasionado que las instituciones de crédito limiten el acceso al financiamiento a personas con características muy definidas, o en su defecto, lo ofrezcan a precios muy altos para cubrir el posible riesgo de crédito que grupos estadísticamente similares ha mostrado previamente, generando obstáculos para la inclusión financiera.

En este contexto, el estudio de las causas y retos de la inclusión financiera ha abierto un nuevo panorama de los factores que podrían incidir en el desempeño financiero de la población, en donde los niveles de alfabetización financiera han comenzado a jugar un papel cada vez más relevante, y han impulsado el desarrollo de programas de educación financiera por parte de las autoridades financieras, ya que, la investigación reciente ha dado evidencia de que los comportamientos, hábitos y conocimientos que conforman el nivel de alfabetización financiera podrían mejorar el desempeño financiero y reducir el riesgo de impago (Hvidberg, 2022; Fujiki, 2020; Urban et al., 2020; Brown et al., 2016; Bruhn et al., 2014, 2016; Mungaray et al., 2021). En particular, la evidencia sugiere que los programas de educación financiera enfocados en mejorar el conocimiento y la conciencia financiera de los individuos pueden mejorar el desempeño de las personas usuarias de servicios financieros e impulsar la penetración del mercado de crédito (Kaiser et al., 2022; Kaiser & Menkhoff, 2017, 2020). De esta forma, los hallazgos de la literatura reciente sugieren que considerar el nivel de alfabetización financiera de las personas en el análisis crediticio de los solicitantes de financiamiento podría reducir la asimetría de información que enfrentan las instituciones financieras.

De esta forma, en este trabajo se plantean dos objetivos principales. El primero, es analizar si el nivel de alfabetización financiera de la población incide en la probabilidad de impago de sus obligaciones con instituciones financieras. Para esto, se utiliza la información de los últimos dos levantamientos de la ENIF (2018 y 2021), ya que, esta encuesta permite realizar mediciones de las capacidades y conocimientos económico-financieros de la población, y al mismo tiempo provee de información sociodemográfica y cierto historial financiero/crediticio de la persona encuestada, variables que son comúnmente usadas por las instituciones de crédito en la actualidad para el análisis crediticio de los solicitantes. El segundo objetivo es analizar si la información referente al nivel de alfabetización financiera de las personas, que incluye sus conocimientos, comportamientos y actitudes financieras (Comisión Nacional Bancaria y de Valores, 2019) podría ser útil para realizar un análisis crediticio. Para este objetivo, además de la metodología estándar

utilizada y considerada en el marco regulatorio actual (regresión logística), se evalúa la robustez de los resultados contrastándolos con modelos de aprendizaje automático.

Los resultados sugieren que existe una relación negativa entre los niveles de alfabetización financiera y la probabilidad de impago, lo cual es consistente con los trabajos de Brown et al. (2016) y Urban et al. (2020). Sin embargo, a diferencia de los trabajos previos, los resultados de este estudio permiten desagregar más el análisis y apuntan a que los beneficios encontrados pueden atribuirse al componente de comportamientos financieros y no necesariamente al nivel de conocimiento financiero que el individuo posee. Adicionalmente, se encuentra que la información sobre los comportamientos financieros tiene un poder explicativo similar al que muestran las variables utilizadas actualmente para explicar el incumplimiento de pago en los créditos (información socioeconómica y de historial financiero/crediticio del individuo), lo que sugiere que incorporar la información de alfabetización financiera del solicitante de crédito podría mejorar la precisión con la que se mide la probabilidad de incumplimiento y reducir el riesgo de crédito.

La relevancia de estos hallazgos podría clasificarse en dos rubros, primordialmente. Primero, la evidencia de que los comportamientos financieros son más relevantes que los conocimientos para el desempeño de los acreditados, sugiere que es necesario valorar la inclusión de mecanismos que los promuevan en el diseño de los programas de educación financiera. En particular, fomentar el desarrollo de comportamientos financieros como la elaboración de presupuestos, el establecimiento de metas financieras, el manejo adecuado de deudas, la cultura del ahorro, entre otros, podría aumentar el alcance de los programas de educación financiera. Segundo, la evidencia muestra que las variables de comportamientos financieros, y no exclusivamente las variables de historial crediticio pueden dar una buena señal del comportamiento futuro de los acreditados. Lo cual podría ser un elemento clave para reducir las asimetrías de información de personas que no cuentan con historial crediticio, permitiéndoles superar algunas de las barreras de acceso al mercado.

El resto del trabajo está estructurado de la siguiente manera: en la siguiente sección se presenta una breve reseña de la literatura que ha estudiado los efectos que la educación financiera tiene sobre el desempeño financiero de las personas. En la tercera sección se detallan las características que hacen que la ENIF sea una fuente de datos de suma utilidad para el análisis de los niveles de alfabetización financiera y su relación con la probabilidad de incumplimiento de sus obligaciones financieras. La cuarta sección muestra los resultados de los factores que inciden en la probabilidad

de impago de la población, poniendo atención en los componentes del nivel de alfabetización financiera. En la quinta sección se presenta un análisis predictivo que contrasta los resultados de la regresión logística con los modelos de aprendizaje automático para medir el poder predictivo de los comportamientos y conocimientos financieros en el análisis crediticio de los individuos. Finalmente, la sexta sección concluye con algunas consideraciones finales sobre los hallazgos del trabajo e intenta abonar con propuestas que se espera puedan contribuir con el diseño de los programas de educación financiera en México.

II. Literatura sobre alfabetización financiera y sus efectos en el desempeño financiero de las personas

Recientemente, ha comenzado a crecer la literatura acerca de cómo la educación financiera afecta ciertos comportamientos financieros e incluso puede dar una señal del desempeño futuro de las personas en el mercado crediticio. Por ejemplo, se ha mostrado que el ser admitido en carreras universitarias relacionadas al área de negocios y economía reduce significativamente el riesgo de impago comparado con los estudiantes admitidos en cualquier otra área (Hvidberg, 2022). Adicionalmente, las personas con un nivel de alfabetización financiera alto tienden a tener más activos financieros que dinero en efectivo, ocasionando un mayor uso de los servicios financieros (Fujiki, 2020). Además, se ha encontrado que, en los estados de Estados Unidos donde se implementaron políticas de educación financiera, las personas tuvieron un incremento en su puntaje crediticio y un decremento en la tasa de morosidad respecto a los estados donde no se implementaron (Urban et al., 2020). En ese mismo país, se encuentra que los jóvenes adultos con más formación financiera y matemática poseen una tasa de morosidad más baja, sin embargo, estos comportamientos tienden a disiparse con la edad (Brown et al., 2016).

Estos resultados parecen observarse también en economías en desarrollo. En Brasil, se ha comprobado que incluir educación financiera a nivel secundaria incrementa el conocimiento financiero y, a su vez, mejora comportamientos financieros, como el hábito del ahorro, (Bruhn et al., 2016). Por su parte, en México, también se ha encontrado una relación entre la alfabetización financiera de la población y su desempeño en el sistema financiero. Por una parte, los cursos de educación financiera que se han implementado en la Ciudad de México que, abordando temas de ahorro, jubilación y uso del crédito, propiciaron un incremento en el índice de conocimiento financiero y en el índice de resultados de ahorro (Bruhn et al., 2014). Asimismo, se ha documentado la existencia de una relación entre el nivel de alfabetización financiera de las personas y sus ingresos, pero a su vez,

elevar el nivel de alfabetización financiera de los mexicanos disminuye su nivel de endeudamiento (Mungaray et al., 2021).

En general, la evidencia sugiere que cuando los programas de educación financiera tienen como objetivo fundamental incrementar el conocimiento y la conciencia financiera de los individuos, estos mejoran el desempeño de las personas usuarias de servicios financieros, y promueven la participación de estas en el sistema financiero. (Kaiser et al., 2022; Kaiser & Menkhoff, 2017, 2020). Además, se encuentra evidencia de efectos positivos indirectos en el comportamiento financiero de los involucrados, aunque de menor magnitud. Destaca que los análisis también concluyen que es más fácil modificar comportamientos sobre el ahorro y más difícil modificar aquellos relacionados con el manejo de deudas.

De esta forma, los hallazgos de la literatura reciente sugieren que el análisis crediticio que realizan las instituciones financieras podría beneficiarse de incluir variables que contengan información acerca del nivel de alfabetización financiera de los solicitantes. Esta información podría tener el potencial de reducir las brechas de asimetría de información que enfrentan las instituciones financieras, al mejorar la clasificación de las personas solicitantes, y dando como resultado una evaluación más precisa del riesgo de crédito. Lo cual podría ser relevante para la inclusión en el mercado de crédito formal de la población que no cuenta con historial crediticio.

III. La ENIF en la medición de la alfabetización financiera y el puntaje crediticio

En esta sección se detalla la fuente de datos utilizada: la Encuesta Nacional de Inclusión Financiera 2018 y 2021, la cual, es la fuente empleada para estimar el índice de alfabetización financiera en México (Comisión Nacional Bancaria y de Valores, 2019). Además, se hace un recuento de los datos que comúnmente se utilizan para los modelos de puntaje crediticio (*credit scoring*) y, posteriormente, se presenta un breve análisis de estadística descriptiva que permite dar un primer acercamiento de la relación entre el nivel de alfabetización financiera de una persona y su riesgo de incumplimiento de pago.

Desde 2012, la CNBV, en conjunto con el Instituto Nacional de Estadística y Geografía (INEGI), realiza la ENIF de forma trienal. Esta encuesta recopila información sobre el acceso y uso de los servicios financieros por parte de la población adulta en México, y desde el levantamiento de 2018 se incluyó información para la evaluación del nivel de alfabetización financiera en México. Esta base de datos permite relacionar el nivel de alfabetización financiera de los individuos con su uso de productos y servicios financieros,

con el comportamiento que las propias personas reportan haber tenido en el mercado crediticio.

Esta encuesta es representativa de la población mexicana adulta¹ y a nivel regional, por tamaño de localidad y por sexo. A partir de esta base de datos, se seleccionaron a aquellos individuos que reportaron contar con algún crédito formal.² Con esta segmentación, se obtiene una muestra de 4,111 observaciones para 2018 y 4,375 para 2021, las cuales representan a poco más de 24.6 millones de personas en 2018 y 27 millones de personas para 2021. Los análisis de esta investigación se realizan tanto sobre el agregado de datos de ambos levantamientos como para cada una de las encuestas, para cual se sigue la propuesta de Pitblado (2009), para asegurar la comparabilidad de la muestra y las variables estudiadas en este documento. De igual forma, los factores de expansión fueron ponderados de tal manera que la suma de estos diera como resultado el total de personas representadas en la edición más reciente de la encuesta, es decir, más de 27 millones de mexicanos y mexicanas entre los 18 y 70 años con crédito formal. A continuación, se muestran las variables seleccionadas que se incorporan tanto en el análisis de puntaje crediticio como en los determinantes de la probabilidad de impago.

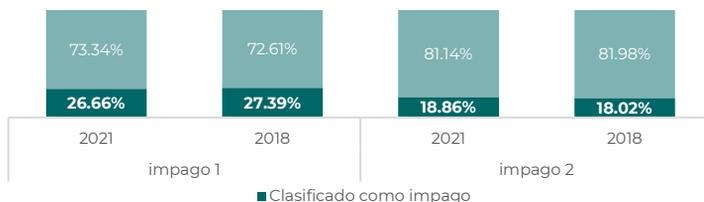
Algo que resulta fundamental para este trabajo es que, la ENIF también recopila información sobre si las personas se han atrasado en alguno de los créditos formales con los que cuenta, lo cual permite discriminar entre aquellas personas que han caído en impago de aquellas que no.³ Así se puede construir una variable que indique si el individuo reporta haber caído en impago o no haberlo hecho. Para robustecer los resultados se consideran dos definiciones diferentes de impago. Para la primera se consideró como impago a aquellas personas que hayan tenido al menos un atraso en alguno de los diferentes tipos de crédito formal con los que contaban (definición uno de impago). Adicionalmente, se consideró como caer en impago a aquellas personas que hayan tenido atrasos más de una vez en al menos un tipo de crédito formal (definición dos de impago). Los cálculos de estas variables muestran consistencia entre ambos levantamientos a pesar de la crisis sanitaria mundial originada por la COVID-19 y las restricciones económicas que esta originó (Gráfica 1).

¹ Para 2018, la ENIF es representativa de la población adulta de entre 18 y 70 años, por lo que el análisis se restringió a esta población en la muestra de 2021.

² Según la definición de Comisión Nacional Bancaria y de Valores, (2021). Las preguntas que indagan acerca del financiamiento en México permiten saber si las personas solicitaron algún tipo de préstamo mediante algún mecanismo informal (caja de ahorro del trabajo o conocidos, casa de empeño, amigos o conocidos o familiares), así como si recurrieron a algún instrumento formal (crédito personal, automotriz, de nómina, de vivienda con Infonavit, Fovissste u otra institución financiera, tarjeta de crédito bancaria, departamental, de tienda de autoservicio, crédito grupal, comunal o solidario).

³ Se eliminaron a aquellos acreditados que respondieron "no sé" o "no responde" sobre si se habían atrasado en los pagos de sus créditos, dejando un total de 4,079 observaciones para 2018 y 4,347 para 2021.

Gráfica 1. Distribución de personas etiquetadas como impago en la población adulta con crédito formal, por período y definición de impago



Fuente: Cálculos propios con información de la ENIF.

Alfabetización financiera en México

En México, el organismo encargado de la educación financiera en el país es el Comité de Educación Financiera (CEF).⁴ En 2017 el CEF publicó la Estrategia Nacional de Educación Financiera (ENEF), en la cual se definió la Educación Financiera como el conjunto de acciones necesarias para que la población adquiera aptitudes, habilidades y conocimientos que le permitan administrar y planear sus finanzas personales, así como usar de manera óptima los productos y servicios que ofrece el sistema financiero en beneficio de sus intereses personales, familiares, laborales, profesionales, y de su negocio.⁵ Posteriormente, la Política Nacional de Inclusión Financiera incluyó la ENEF como uno de sus ejes.

En este contexto, se ha definido a la alfabetización financiera como el conjunto de conocimientos, habilidades y aptitudes que permiten alcanzar un bienestar económico (Comisión Nacional Bancaria y de Valores, 2019). Así, la alfabetización se refiere a la combinación de la concientización, conocimientos, habilidades y comportamientos necesarios para tomar decisiones financieras sólidas y eventualmente lograr el bienestar financiero individual, se mide a través del índice de alfabetización financiera.⁶ Este índice se compone de tres diferentes subíndices, el subíndice de conocimientos

⁴ Este comité está conformado por la Secretaría de Hacienda y Crédito Público (SHCP), la Secretaría de Educación Pública (SEP), la Comisión Nacional para la Protección y Defensa de los Usuarios de Servicios Financieros (CONDUSEF), la Comisión Nacional Bancaria y de Valores (CNBV), la Comisión Nacional de Seguros y Fianzas (CNSF), la Comisión Nacional del Sistema de Ahorro para el Retiro (CONSAAR), el Instituto para la Protección al Ahorro Bancario (IPAB) y el Banco de México (BANXICO).

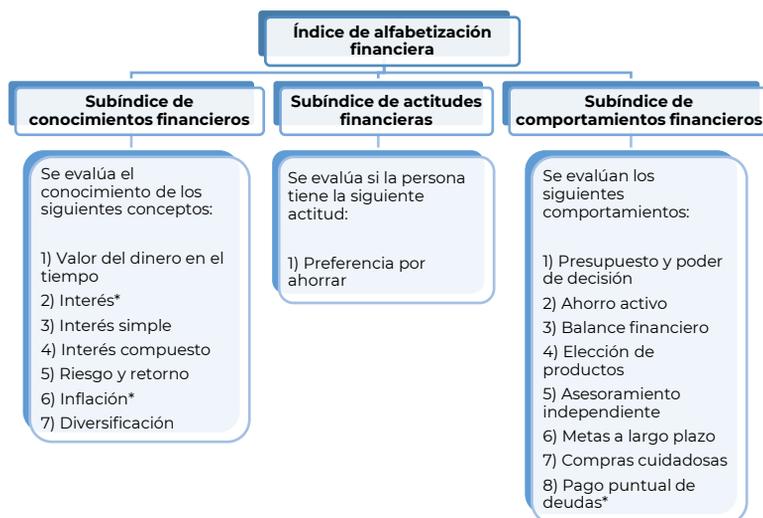
⁵ La Organización para la Cooperación y el Desarrollo Económicos (OCDE) define la educación financiera como "el proceso por el cual los consumidores/inversionistas financieros mejoran su comprensión de los productos financieros, los conceptos y los riesgos y, a través de información, instrucción y/o el asesoramiento objetivo desarrollan las habilidades y confianza para ser más conscientes de los riesgos y oportunidades financieras, tomar decisiones informadas, saber a dónde ir para obtener ayuda y ejercer cualquier acción eficaz para mejorar su bienestar económico."

⁶ En 2021 cambió el nombre a índice de competencias económico-financieras.

financieros, el subíndice de comportamientos financieros y el de actitudes financieras (Figura 1).⁷

Para este análisis, las variables de alfabetización financiera se dividen en dos grupos. El primero, representa las variables de conocimientos y actitudes financieras y está conformado por 6 variables extraídas de los conceptos que componen el subíndice de conocimientos financieros más una variable, la cual es utilizada para la medición del subíndice de actitudes financieras.⁸ El segundo, representa la parte de los comportamientos financieros y se conforma por las variables que componen este subíndice.⁹

Figura 1. Estructura del índice de alfabetización financiera



Nota: Para los análisis posteriores, los conocimientos 1) y 6) fueron reemplazados por el conocimiento de ambos conceptos, esto debido a que más del 94% de los adultos con crédito formal reportó el conocer estos conceptos por separado. De igual manera, el comportamiento financiero 8) fue eliminado de los análisis debido a su alta correlación con las variables de impago.

Fuente: Elaboración propia con información a partir de Alfabetización Financiera en México (CNBV, 2019)

⁷ En el levantamiento de la ENIF 2021 se modificó levemente la manera de calcular este índice, haciendo imposible el replicar la nueva metodología para los datos de la ENIF 2018, y aunque es posible hacer el análisis utilizando metodologías diferentes para los respectivos años de las encuestas, la adición de un nuevo comportamiento financiero para el año 2021 hace que sea más viable utilizar la metodología de medición de alfabetización financiera de 2018 para ambos levantamientos de la encuesta.

⁸ Debido al poco porcentaje de la población que no conocía los conceptos de interés e inflación, se unificaron estos conceptos en un nuevo conocimiento financiero llamado "Interés e Inflación".

⁹ Sin embargo, dentro del subíndice de comportamientos financieros existe un comportamiento que está altamente correlacionado con las variables objetivo, ya que, dicho comportamiento evalúa si la persona considera realizar el pago puntual de sus deudas. Para evitar sesgos en la estimación debido a la alta correlación se eliminó esta variable, haciendo que el grupo de comportamientos financieros quede representado por 7 variables en total.

La ENIF en el puntaje crediticio

La información que se incluye en los modelos de clasificación para medir el riesgo de incumplimiento varía dependiendo de si el individuo evaluado es una persona, una pequeña o mediana empresa (PyME) o un corporativo, o bien, dependiendo el tipo de financiamiento que se pretende obtener. Para el caso de personas, las variables más usualmente utilizadas tanto en los trabajos empíricos como en el marco regulatorio son las variables socioeconómicas y las de historial financiero/crediticio. Por ejemplo, Glennon et al. (2008) utilizan información de tres grandes burós de crédito estadounidenses para alimentar modelos de puntaje crediticio que les permitan predecir la probabilidad de caer en impago de una persona; Greene (1992) utiliza nueve variables socioeconómicas combinadas con cuatro variables de historial financiero con un objetivo similar; Alonso & Carbó, (2020) y Huang et al. (2007) hacen uso de una combinación de variables socioeconómicas y de historial financiero para llevar a cabo sus predicciones, y concluyen que las variables financieras aportan información valiosa para el análisis.

En México, el marco regulatorio de las instituciones de crédito (circular única de bancos) solicita que para calcular la probabilidad de incumplimiento de un acreditado se utilice la regresión logística y que se consideren variables relacionadas con el historial crediticio y las características sociodemográficas asociadas al individuo evaluado. Estas incluyen sexo, tipo de crédito, porcentaje de la deuda que hace falta por cubrir, monto por pagar, antigüedad del acreditado en la entidad, número de atrasos previos, meses transcurridos desde el último atraso, máximo número de atrasos presentado en los últimos cuatro meses, entre otras, las cuales varían por tipo de crédito solicitado.

Adicionalmente, los bancos requieren información de las personas que solicitan un nuevo crédito, entre lo que suele considerarse datos sociodemográficos y financieros que permitan evaluar si otorgan o no un crédito. Las instituciones deben comunicar parte de esta información a la CNBV como muestra de que han realizado una evaluación adecuada de sus nuevos clientes. Esta se compone de datos como sexo, edad, estado civil, nombre, Registro Federal de Contribuyente (RFC), Clave Única de Registro de Población (CURP), municipio donde labora, estado donde labora, si cuenta con trabajo formal, así como si cuenta con garantía u otra información que podría incidir en el comportamiento del pago de las obligaciones.

Si bien, el volumen de información solicitado es considerable, hasta donde se tiene conocimiento, en ningún caso se considera información acerca de los conocimientos y comportamientos financieros de las personas, lo cual, podría deberse tanto a la escasez de dichos indicadores como a que esta corriente de estudio tiene poca antigüedad. De esta forma, la ENIF

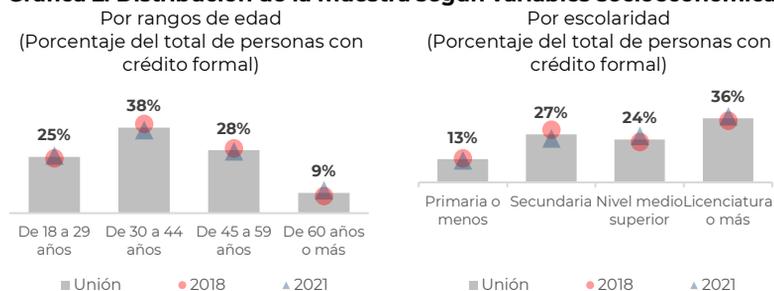
representa una fuente de información que conjunta variables sociodemográficas, financieras y de historial crediticio con indicadores sobre el nivel de educación financiera de la población. Por lo que, hasta donde se tiene conocimiento, este documento representa el primer análisis con representatividad a nivel nacional que se realiza sobre el efecto de la alfabetización financiera en la probabilidad de incumplimiento de la población.

Variables adicionales utilizadas

Variables socioeconómicas

En primer lugar, la ENIF contiene una gran diversidad de variables socioeconómicas de la persona entrevistada. Entre las que se incluye la edad, la escolaridad, el sexo, el estatus laboral y la situación conyugal, las cuales han sido utilizadas para el análisis de la probabilidad de impago realizado en este trabajo. La variable de ingreso fue dejada fuera del análisis debido a su alta tasa de no respuesta. Sin embargo, características como el estatus laboral, región en la que habita, escolaridad y tipo de localidad contienen información relevante que podrían operar como *proxys* de esta variable. En la Gráfica 2 se presenta la distribución de algunas de las variables socioeconómicas utilizadas para el análisis¹⁰. Se aprecia que la distribución de esta información no presenta diferencias significativas entre los levantamientos de 2018 y 2021, lo cual nos permite hacer inferencias estadísticas en el agregado de datos (Unión) y extrapolar los resultados (Pitblado, 2009).

Gráfica 2. Distribución de la muestra según variables socioeconómicas



Nota: Las gráficas de la distribución de todas las variables socioeconómicas utilizadas para el análisis se pueden obtener a petición de los autores.

Fuente: Cálculos propios con información de la ENIF.

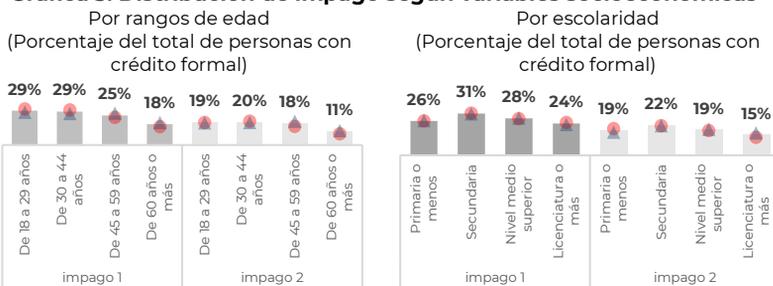
En particular, la mayor parte de la población con crédito formal está en un rango de edad de entre los 30 y 44 años. De igual manera, el acceso al financiamiento formal permea más a medida que la escolaridad avanza, ya que el 36% de la población adulta que cuenta con financiamiento tiene el

¹⁰ Lo observado se mantiene para el resto de las variables, sexo, localidad, situación conyugal, estatus laboral y región.

grado de licenciatura o más, comparado con el 27% y 24% de la población con estudios hasta secundaria y nivel medio superior, respectivamente. La brecha en el acceso al crédito entre las personas que viven en comunidades rurales (definidas en este trabajo como poblaciones con hasta 15 mil habitantes) en relación con quienes viven en comunidades urbanas no disminuyó de 2018 a 2021, ya que el 75% de la población con crédito formal vive en comunidades urbanas y solo el 25% en comunidades rurales. Además, 79% de la población adulta con crédito formal posee un trabajo remunerado y solo el 21% de los que tienen crédito no están percibiendo ingresos, esto podría deberse a que las entidades financieras catalogan como más riesgoso prestar dinero a una persona que no recibe ingresos, lo cual se traduce en incapacidad de pago. Aunque, por la naturaleza de la encuesta no es posible determinar si el estado laboral pudo cambiar después de haber adquirido el financiamiento.

En cuanto a la relación de las variables socioeconómicas con su propensión a caer en impago destaca que en cada uno de los grupos existe cierto nivel de heterogeneidad, lo cual sugiere que podrían aportar información importante para su puntaje crediticio. En particular, la edad y el nivel de escolaridad del acreditado parecieran tener una relación negativa con su probabilidad de incumplimiento (Gráfica 3). Con ambas definiciones de impago, se observa que a partir de los 30 años la tasa de acreditados que incumplen sus créditos disminuye conforme la edad aumenta. Respecto al nivel educativo, después de haber estudiado la secundaria, la tasa de incumplimientos disminuye. Asimismo, se observa un mayor incumplimiento por parte de la población que vive en localidades urbanas. Lo último es de gran relevancia en tema de inclusión, ya que solo el 25% de la población con crédito formal vive en localidades rurales. La brecha existente podría no deberse exclusivamente a temas de morosidad, sino que varios factores, incluidos la falta de historial financiero, pueden ser los contribuyentes a dicha brecha. Sin embargo, la estadística descriptiva dista de ser concluyente, por lo que, es necesario profundizar en el análisis.

Gráfica 3. Distribución de impago según variables socioeconómicas



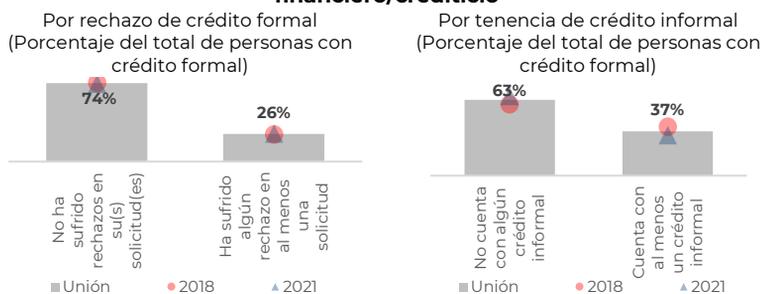
Nota: Las gráficas de la distribución de impago de todas las variables socioeconómicas utilizadas para el análisis se pueden obtener a petición de los autores.

Fuente: Cálculos propios con información de la ENIF.

Variables de historial financiero/crediticio

Como se había señalado, una de las ventajas que ofrece la ENIF sobre otras bases de datos es que permite obtener información que podría dar referencia del historial crediticio de las personas. Para este trabajo, se seleccionaron 6 características del entrevistado que contienen información de su historial financiero/crediticio. Igual que con las variables socioeconómicas, existe poca diferencia en la proporción de población adulta con crédito formal entre el período de 2018 y 2021¹¹. El 26% de la población con crédito formal ha sufrido rechazo en alguna de sus solicitudes de crédito formal y el 37% de la muestra cuenta con algún crédito informal (Gráfica 4).

Gráfica 4. Distribución de la muestra según variables de historial financiero/crediticio



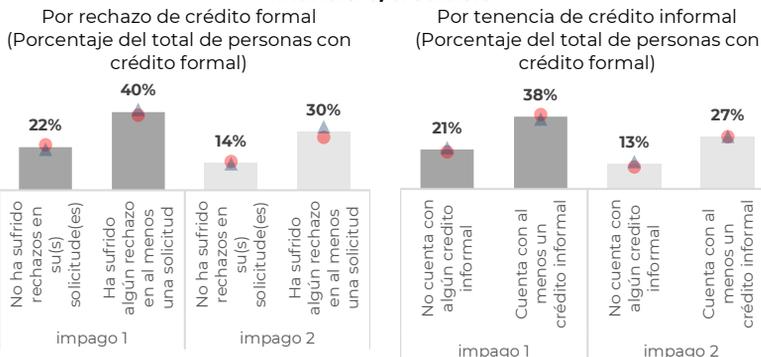
Nota: Las gráficas de la distribución de todas las variables de historial financiero/crediticio utilizadas para el análisis se pueden obtener a petición de los autores.

Fuente: Cálculos propios con información de la ENIF.

Aquellas personas que están más expuestas al mercado crediticio, ya sea formal o informal, tienden a incumplir más en los pagos de sus deudas (Gráfica 5). Para el caso de personas que cuentan con algún tipo de crédito informal, se observa que independientemente de la definición de impago utilizada, éstos son los que tienen una propensión mayor a caer en incumplimiento.

¹¹ Lo observado se mantiene para el resto de las variables, cuenta de apoyo, cuenta de inversión, cantidad de cuentas de captación y cantidad de créditos informales.

Gráfica 5. Distribución de impago según variables de historial financiero/crediticio

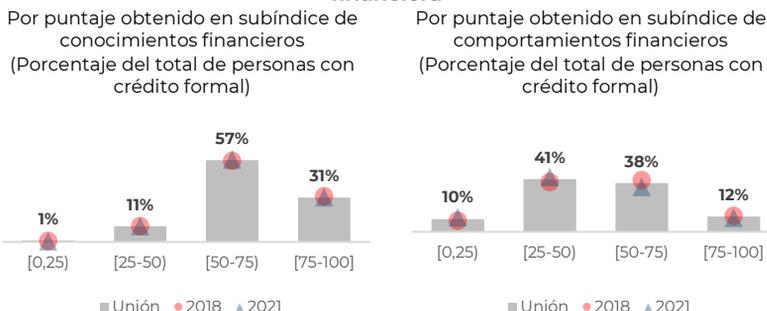


Nota: Las gráficas de la distribución de impago de todas las variables de historial financiero/crediticio utilizadas para el análisis se pueden obtener a petición de los autores.
Fuente: Cálculos propios con información de la ENIF.

Variables de alfabetización financiera

Finalmente, se utiliza el índice de alfabetización financiera, dividido en dos subgrupos diferentes: el primero hace referencia a los conocimientos y las actitudes financieras que posee la persona encuestada y el otro a sus comportamientos financieros (Figura 1). Al igual que en los otros grupos de variables, se observa consistencia entre ambos levantamientos (Gráfica 6). Aproximadamente 6 de cada 10 personas tienen un puntaje de entre 50 y 75 para el índice de alfabetización financiera. Para el caso del subíndice de conocimientos financieros, casi el 90% de la muestra está por encima de los 50 puntos, en el caso del de comportamientos financieros solo el 50% se encuentra por encima de dicho puntaje, lo cual es un indicador de que es posible realizar el análisis con los datos agregados.

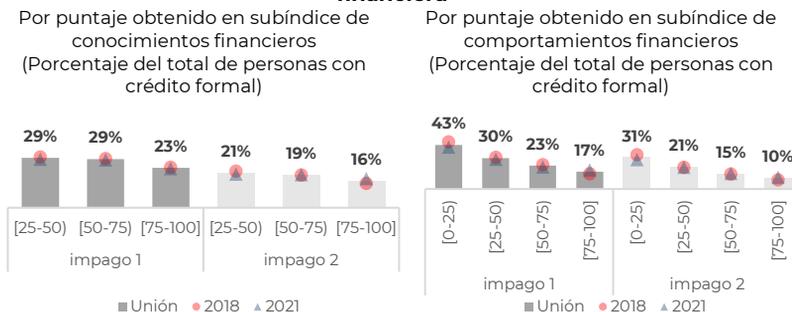
Gráfica 6. Distribución de la muestra según variables de alfabetización financiera



Nota: Las gráficas de la distribución de todas las variables de alfabetización financiera utilizadas para el análisis se pueden obtener a petición de los autores.
Fuente: Cálculos propios con información de la ENIF.

La heterogeneidad en la distribución de impago (ver Gráfica 7) dentro de este grupo sugiere que esta información puede ser relevante para medir el riesgo de crédito. Se observa que a mayor nivel de alfabetización financiera el porcentaje de personas que cae en impago disminuye, evidenciando una relación negativa. Sin embargo, pareciera que dicha tendencia es explicada, en mayor medida, por los comportamientos financieros de las personas evaluadas. Ya que se observa una caída de aproximadamente 25 puntos porcentuales entre el grupo con puntajes de entre 0 y 25 en el subíndice de comportamientos (42.85% de personas caen en impago, usando la definición uno), respecto a los que tienen un puntaje de entre 75 y 100 (16.74% de personas caen en impago). Dicha tendencia también se observa, aunque en menor medida, en el subíndice de conocimientos financieros.

Gráfica 7. Distribución de impago según variables de alfabetización financiera



Nota: Las gráficas de la distribución de impago de todas las variables de alfabetización financiera utilizadas para el análisis se pueden obtener a petición de los autores.

Fuente: Cálculos propios con información de la ENIF.

Es importante señalar que, por la naturaleza de los datos, una posible debilidad del estudio es no tener certeza sobre la temporalidad de cuando se tuvo acceso al crédito y el período en que se adquirieron las competencias económico-financieras. Sin embargo, es de esperar que el nivel de alfabetización financiera se definiera antes y no se modifique en el corto plazo.¹² Además, aunque en la definición de impago utilizada no es posible determinar el período preciso de atraso en el pago de los créditos, la información se refiere a un período reciente (últimos 12 meses), lo que permite suponer que este se dio posterior a que el individuo adquiriera su nivel actual de alfabetización financiera.

¹² El índice de alfabetización financiera de la población no presentó cambios entre 2018 y 2021 (Comisión Nacional Bancaria y de Valores, 2022).

IV. Alfabetización financiera y probabilidad de impago

En México, las instituciones de crédito que captan recursos del público están obligadas a contar con metodologías para medir y gestionar el riesgo de crédito, las cuales se sustentan en las recomendaciones del Comité de Basilea.¹³ Este marco normativo ofrece la posibilidad de utilizar metodologías basadas en técnicas estadísticas tradicionales, como es el caso de la regresión logística, pero es flexible para permitir a las instituciones que desarrollen sus propias metodologías internas para la medición y cálculo de las reservas financieras que deberán constituir para mitigar las pérdidas potenciales por el riesgo de crédito incurrido.¹⁴

En esta sección se presenta un análisis de los determinantes de la probabilidad de impago de la población, a partir de la información de la ENIF 2018 y 2021. En particular, se presentan una serie de estimaciones considerando tanto de forma agregada los subíndices que conforman el índice de alfabetización financiera, explicados en la sección III de este documento, como de manera desagregada, lo que permite identificar los componentes que están incidiendo de forma significativa en el desempeño financiero de las personas. Además, dada la robusta consistencia entre los levantamientos de la ENIF que se mostró en la sección anterior, se presentan los análisis de forma agregada.¹⁵

Modelos de puntaje crediticio

Los modelos de puntaje crediticio como métodos para evaluar el riesgo de crédito aparecieron desde los años 70's, pero no fue hasta los años 90's que se generalizó su uso gracias al avance computacional. Estos modelos asignan a cada persona, o empresa evaluada, una calificación o clasificación en función de sus características conocidas. La calificación intenta reflejar su riesgo de incumplimiento, por lo que es común que este problema de evaluación de riesgo se convierta en un problema de clasificación binaria. En este sentido, el problema se resume en asignar a un individuo una de dos categorías posibles: "bueno" o "malo", "aceptado" o "rechazado", "default" o "no default", etc. Lo

¹³ "El Comité de Supervisión Bancaria de Basilea es un comité de autoridades de supervisión bancaria que fue establecido por los gobernadores de los bancos centrales de los países del Grupo de los Diez en 1975. Está formado por altos representantes de las autoridades de supervisión bancaria y los bancos de sus 12 países miembros (G10, Luxemburgo y España)", (Bank for International Settlements, 2006).

¹⁴ El detalle del marco regulatorio que rige a las instituciones de crédito puede encontrarse en las disposiciones de carácter general aplicables a las instituciones de crédito (circular única de bancos), que puede consultarse en la siguiente liga: <https://www.cnbv.gob.mx/Normatividad/Disposiciones%20de%20car%C3%A1cter%20general%20aplicables%20a%20las%20instituciones%20de%20cr%C3%A9dito.pdf>

¹⁵ Farhat & Robb (2014); Pitblado (2009) y Heeringa, West & Berglund (2017), dan una explicación del tratamiento por el cual se procesan los datos provenientes de encuestas, en particular, Pitblado, 2009 da un análisis y recetario para el procesamiento de un agregado de datos conformado por dos encuestas levantadas en diferentes períodos. El análisis estadístico mostrado a continuación es producto del procesamiento de datos comentado anteriormente.

anterior, permite decidir si otorgar o no un préstamo al solicitante o cuantificar el riesgo de que un acreditado deje de cumplir con sus obligaciones.

Dentro de los modelos de puntaje crediticio existen diversas metodologías que se utilizan para evaluar el riesgo de crédito. Los modelos más comunes están basados en métodos estadísticos, de los cuales destaca la regresión logística (Dastile et al., 2020), la cual no solo es popular por su fácil implementación sino por la intuición e interpretación que ofrecen sus resultados. Por ello, este es el método que se utiliza en esta sección.

No obstante, en las últimas décadas el sector financiero comenzó a adoptar nuevas técnicas para la evaluación del riesgo de crédito. Por ejemplo, Petropoulos et al. (2018) utilizan los modelos de deep learning y extreme gradient boost (XGBoost) para predecir la probabilidad de impago de un deudor, mientras que Alonso & Carbó, (2020) proponen una medición de los beneficios y los costos de implementar métodos de aprendizaje automático para la medición de riesgo de crédito mediante este tipo de metodologías. Los autores señalan que estos nuevos métodos pueden ser más precisos que los métodos estadísticos comúnmente empleados. Sin embargo, interpretar los parámetros de estos modelos es una tarea sumamente difícil, haciendo complejo explicar de manera precisa porqué el acreditado podría caer en incumplimiento.

Para un marco más general, Abdou & Pointon (2011); Assef & Steiner (2020); Dastile et al. (2020) y Louzada et al. (2016) presentan revisiones de la literatura acerca de las técnicas que han sido utilizadas para la elaboración de modelos de puntaje crediticio en los últimos años. Si bien, el uso de algoritmos de aprendizaje automático ha ido en aumento, actualmente el método más utilizado es la regresión logística. A pesar de ello, la diversidad de métodos de aprendizaje automático abre la posibilidad de no estar limitados a una sola medición y permite contrastar la robustez de los resultados (Dastile et al., 2020).¹⁶

Como se había planteado, este trabajo busca dos objetivos principales. En primer lugar, se evalúa si los componentes de la alfabetización financiera de la población inciden en la probabilidad de impago de sus obligaciones con instituciones financieras. Esto se realizará mediante la metodología de regresión logística en esta sección. En segundo lugar, se analiza si la información referente al conocimiento, comportamiento y las actitudes financieras aportan información relevante sobre el desempeño financiero de la población. Para este objetivo, además de la metodología

¹⁶ Entre estos están las Redes neuronales, K-vecinos más cercanos, AdaBoost representando al Boosting y Bosques Aleatorios representando al Bagging. Estas metodologías serán utilizadas en la sección V.

estándar (regresión logística) se contrastan los resultados con modelos de aprendizaje automático, lo que permitirá evaluar la robustez de los resultados.

Efecto del nivel de alfabetización financiera

Para cada uno de los levantamientos se estimaron regresiones con diferentes especificaciones, las cuales incluyen, en primera instancia, solo información de los niveles del nivel de alfabetización financiera de la población y, posteriormente, se controla por sus características socioeconómicas y por su historial financiero\crediticio, con el fin de verificar la robustez de los resultados obtenidos.

Los resultados para las estimaciones con la muestra agregada y la definición uno de impago muestran que, ante todas las especificaciones consideradas, el subíndice de comportamientos financieros parece incidir en la probabilidad de impago de los acreditados (Tabla 1).¹⁷ En particular, para el modelo de puntaje crediticio alimentado con las variables socioeconómicas y de historial financiero, se observa que las variables de alfabetización financiera reducen la probabilidad de que un acreditado caiga en impago. En particular, aumentar en una desviación estándar el subíndice de comportamientos financieros de un acreditado reduce su probabilidad de caer en impago en aproximadamente 6 puntos porcentuales (pp). No obstante, la magnitud de los efectos parciales en la variable de comportamientos financieros decae en la medida en que se le agregan más controles al modelo. Los resultados muestran robustez ante los cambios en las especificaciones. Por otra parte, lo contrario ocurre para las variables de conocimientos y actitudes financieras, las cuales, no muestra tener algún impacto estadísticamente significativo en la probabilidad de incumplimiento.

Adicionalmente, los resultados para las variables de control son en su mayoría en el sentido esperado según el análisis inicial de la sección III. En primer lugar, no parece haber diferencia significativa en la probabilidad de impago entre los jóvenes de entre 18 y 29 años y adultos de entre 30 y 44 años. Sin embargo, a partir de los 45 años la propensión de caer en impago disminuye conforme aumenta la edad.

¹⁷ En el análisis de regresión logística con variables dicotómicas como explicativas de los resultados se interpretan en referencia a aquellos niveles de las variables que quedan fuera de la estimación. Por ejemplo, en el caso de la inclusión de variables socioeconómicas se está comparando con la probabilidad de incumplimiento de hombres, rurales, de 18 a 20 años, casados, con educación primaria o menor, de la región Sur y sin trabajo remunerado.

Tabla 1. Efectos de la alfabetización financiera
(Definición 1 de impago)¹⁸

	(1) EDU <i>impago_1</i>	(2) EDU+SE <i>impago_1</i>	(3) EDU+FIN <i>impago_1</i>	(4) EDU+SE+FIN <i>impago_1</i>
Alfabetización financiera				
Conocimientos financieros	-0.0118	-0.0117	-0.0142*	-0.0114
Actitudes financieras	-0.00840	-0.00753	-0.00667	-0.00592
Comportamientos financieros	-0.0704***	-0.0690***	-0.0618***	-0.0597***
Socioeconómicas				
Edad de 30 a 44		-0.0181		-0.0148
Edad de 45 a 59		-0.0728***		-0.0597***
Edad de 60 o más		-0.157***		-0.117***
Sexo Mujer		-0.0274*		-0.0231*
Localidad Urbana		0.0461***		0.0269*
Escolaridad Secundaria		0.0521**		0.0404**
Escolaridad Media superior		0.0413*		0.0182
Escolaridad Licenciatura o más		0.0506**		0.0162
Situación conyugal Separado		0.0664***		0.0505**
Situación conyugal Soltero		-0.0277		-0.0140
Región Centro Sur y Oriente		-0.0518**		-0.0515**
Región Ciudad de México		-0.115***		-0.0847***
Región Occidente y Bajío		-0.0786***		-0.0574***
Región Noreste		-0.0355*		-0.0268
Región Noroeste		0.0106		0.00753
Estatus lab. Trabajador informal		0.0509**		0.0129
Estatus lab. Trabajador formal		-0.0311		-0.0565***
Historial financiero/crediticio				
Tiene cuenta de apoyo			-0.0252	-0.00946
Tiene cuenta de inversión			0.0160	0.0116
Núm. de cuentas de captación			-0.00918	-0.00180
Rechazo anterior de crédito			0.150***	0.137***
Tiene crédito informal			0.122***	0.109***
Núm. de créditos formales			0.0283***	0.0303***
<i>Número de observaciones:</i>	8,426	8,426	8,426	8,426
<i>Población representada:</i>	27,002,686	27,002,686	27,002,686	27,002,686

Nota: La variación en la probabilidad estimada se calculó para un acreditado hipotético cuyos valores en los regresores son iguales a la media poblacional para el caso de variables continuas, para variables categóricas se calculó el cambio discreto desde el nivel base. Los errores estándar se presentan entre paréntesis debajo de los efectos marginales. La significancia estadística se presenta a niveles de 10%*, del 5%** o del 1%***.

Fuente: Cálculos propios con la información de la ENIF.

De igual manera, la literatura muestra que las mujeres tienden a ser menos morosas que los hombres. Los resultados de este análisis muestran que la diferencia en probabilidades de impago entre un acreditado hombre y una acreditada mujer es de 3 puntos porcentuales, siendo la probabilidad de impago de la mujer acreditada menor que la del hombre. Un resultado que podría requerir análisis adicional es que los acreditados que estudian hasta el

¹⁸ Los resultados de las variables de control muestran robustez ante cambios en la definición de impago, los cuales se pueden consultar en el anexo A.11.

nivel secundaria son los que tienen la mayor probabilidad de caer en impago (incluso más que aquellos que estudian hasta primaria o menos). Para las variables de historial financiero/crediticio se observan los resultados esperados. Los acreditados con mayor exposición al mercado crediticio informal o con evidencia de mayores niveles de endeudamiento, tienen probabilidades de impago más grandes. Además, aquellos acreditados con al menos un rechazo anterior en alguna de sus solicitudes de crédito tienen probabilidades de impago mayores que aquellos que no han sido rechazados, lo cual podría estar reflejando que el análisis crediticio de las instituciones de crédito está logrando segmentar en cierta medida a los acreditados más riesgosos.

Tabla 2. Efectos de la alfabetización financiera por período

(Definición 1 de impago)

	(1) EDU <i>impago_1</i>	(2) EDU+SE <i>impago_1</i>	(3) EDU+FIN <i>impago_1</i>	(4) EDU+SE+FIN <i>impago_1</i>
2018				
Conocimientos financieros	-0.00867	-0.0103	-0.00949	-0.00861
Actitudes financieras	-0.00203	0.00110	-0.00287	-0.000136
Comportamientos financieros	-0.0818***	-0.0795***	-0.0751***	-0.0729***
Número de observaciones:	4,079	4,079	4,079	4,079
Población representada:	24,511,766	24,511,766	24,511,766	24,511,766
2021				
Conocimientos financieros	-0.0143	-0.0156	-0.0183*	-0.0162
Actitudes financieras	-0.0138	-0.0124	-0.00945	-0.00851
Comportamientos financieros	-0.0609***	-0.0584***	-0.0502***	-0.0475***
Número de observaciones:	4,347	4,347	4,347	4,347
Población representada:	27,002,686	27,002,686	27,002,686	27,002,686
Unión				
Conocimientos financieros	-0.0118	-0.0117	-0.0142*	-0.0114
Actitudes financieras	-0.00840	-0.00753	-0.00667	-0.00592
Comportamientos financieros	-0.0704***	-0.0690***	-0.0618***	-0.0597***
Número de observaciones:	8,426	8,426	8,426	8,426
Población representada:	27,002,686	27,002,686	27,002,686	27,002,686
Controles				
Socioeconómicas	No	Si	No	Si
Historial financiero/crediticio	No	No	Si	Si

Nota: La variación en la probabilidad estimada se calculó para un acreditado hipotético cuyos valores en los regresores son iguales a la media poblacional para el caso de variables continuas, para variables categóricas se calculó el cambio discreto desde el nivel base. Los errores estándar se presentan entre paréntesis debajo de los efectos marginales. La significancia estadística se presenta a niveles de 10%*, del 5%*** o del 1%***.

Fuente: Elaboración propia con información de la ENIF 2018 y 2021.

El análisis realizado con las muestras separadas muestra resultados similares a lo encontrado con el análisis agregado. La contribución del componente de comportamiento financiero se mantiene ante cambios en las especificaciones de las estimaciones, en la muestra utilizada o en la definición de impago utilizada. En particular, el efecto del componente de comportamientos financieros en la probabilidad de incumplimiento se mantiene entre 5 y 7 pp por cada desviación estándar que se incrementa, para

la definición uno de impago y de entre 4 y 7 pp para la definición dos. Esto sugiere que promover el desarrollo de buenos comportamientos financieros podría generar menores niveles de morosidad e incumplimiento de pago entre la población acreditada, lo que podría traducirse en mejores condiciones y menor costo del financiamiento (ver Tabla 2 y Tabla 3).

De igual forma, es interesante el hecho de que el subíndice de conocimientos financieros no parece incidir en la probabilidad de impago de la población. Aunque los resultados presentan el signo esperado, para los conocimientos financieros, los resultados no son estadísticamente distintos de cero en la mayoría de las especificaciones realizadas y esto se mantiene ante cambios en la muestra utilizada y en la definición de impago considerada (Tabla 2 y Tabla 3). En relación con las actitudes financieras, los resultados sugieren que al igual que el subíndice de conocimientos financieros, estas no parecen afectar el desempeño de la población en el mercado de crédito formal.

Tabla 3. Efectos de la alfabetización financiera por período

(Definición 2 de impago)

	(1) EDU <i>impago_2</i>	(2) EDU+SE <i>impago_2</i>	(3) EDU+FIN <i>impago_2</i>	(4) EDU+SE+FIN <i>impago_2</i>
2018				
Conocimientos financieros	-0.0167	-0.0128	-0.0146	-0.0109
Actitudes financieras	-0.000748	0.00162	-0.00167	0.000363
Comportamientos financieros	-0.0698***	-0.0627***	-0.0575***	-0.0523***
<i>Número de observaciones:</i>	4,079	4,079	4,079	4,079
<i>Población representada:</i>	24,511,766	24,511,766	24,511,766	24,511,766
2021				
Conocimientos financieros	-0.00219	-0.00528	-0.00754	-0.00693
Actitudes financieras	-0.000864	0.000251	0.00223	0.00292
Comportamientos financieros	-0.0546***	-0.0531***	-0.0418***	-0.0399***
<i>Número de observaciones:</i>	4,347	4,347	4,347	4,347
<i>Población representada:</i>	27,002,686	27,002,686	27,002,686	27,002,686
Unión				
Conocimientos financieros	-0.00917	-0.00784	-0.0104	-0.00753
Actitudes financieras	-0.000650	-0.000297	0.000177	0.000607
Comportamientos financieros	-0.0620***	-0.0583***	-0.0493***	-0.0464***
<i>Número de observaciones:</i>	8,426	8,426	8,426	8,426
<i>Población representada:</i>	27,002,686	27,002,686	27,002,686	27,002,686
Controles				
Socioeconómicas	No	Sí	No	Sí
Historial financiero/crediticio	No	No	Sí	Sí

Nota: La variación en la probabilidad estimada se calculó para un acreditado hipotético cuyos valores en los regresores son iguales a la media poblacional para el caso de variables continuas, para variables categóricas se calculó el cambio discreto desde el nivel base. Los errores estándar se presentan entre paréntesis debajo de los efectos marginales. La significancia estadística se presenta a niveles de 10%*, del 5%** o del 1%***.

Fuente: Cálculos propios con la información de la ENIF.

Efecto de los componentes de la alfabetización financiera

Aunque la evidencia sobre la existencia de una relación negativa entre el subíndice de comportamientos financieros con la probabilidad de caer en impago dentro del mercado crediticio formal parece contundente, analizar qué componentes del índice de alfabetización financiera son los que contribuyen a ello podría contribuir a direccionar los esfuerzos realizados por los encargados de promover y mejorar el acceso al financiamiento, a través de políticas encaminadas a fortalecer la alfabetización financiera de la población.

Tabla 4. Efectos de los componentes de la alfabetización financiera
(Definición 1 de impago)¹⁹

	(1) EDU impago_1	(2) EDU+SE impago_1	(3) EDU+FIN impago_1	(4) EDU+SE+FIN impago_1
Conocimientos financieros				
Valor del dinero en el tiempo	-0.00874	-0.000917	-0.00890	-0.000815
Interés e Inflación	0.00322	0.00476	-0.00660	-0.00289
Interés simple	0.0122	0.00791	0.0132	0.0125
Interés compuesto	-0.0464**	-0.0366*	-0.0506***	-0.0427**
Riesgo y retorno	-0.0236	-0.0275	-0.0203	-0.0226
Diversificación	-0.00124	0.000762	-0.000175	0.00126
Actitudes financieras				
Preferencia por ahorrar	-0.0167	-0.0126	-0.0129	-0.0107
Comportamientos financieros				
Presupuesto y poder de decisión	-0.00208	0.00417	-0.00854	0.000482
Ahorro activo	-0.0204	-0.0323**	-0.0397**	-0.0473***
Balance financiero	-0.193***	-0.185***	-0.126***	-0.125***
Elección de productos	0.000602	0.00604	-0.0129	-0.00842
Asesoramiento independiente	-0.000873	-3.64e-06	-0.00430	-0.00284
Metas a largo plazo	-0.0243*	-0.0300**	-0.0235*	-0.0276**
Compras cuidadosas	-0.108***	-0.0978***	-0.0825***	-0.0764***
Controles				
Socioeconómicas	No	Si	No	Si
Historial financiero/crediticio	No	No	Si	Si
Número de observaciones:	8,426	8,426	8,426	8,426
Población representada:	27,002,686	27,002,686	27,002,686	27,002,686

Nota: La variación en la probabilidad estimada se calculó para un acreditado hipotético cuyos valores en los regresores son iguales a la media poblacional para el caso de variables continuas, para variables categóricas se calculó el cambio discreto desde el nivel base. Los errores estándar se presentan entre paréntesis debajo de los efectos marginales. La significancia estadística se presenta a niveles de 10%*, del 5%** o del 1%***.

Fuente: Cálculos propios con la información de la ENIF.

Las actitudes y los conocimientos financieros que conforman el índice de alfabetización financiera no muestran efectos estadísticamente significativos en la probabilidad de incumplimiento, a excepción del conocimiento sobre el cálculo de interés compuesto. El hecho de que el subíndice agregado de conocimientos financieros no muestre efectos sobre la probabilidad de impago podría explicarse por el hecho de que el resto de

¹⁹ Los resultados son robustos por definición de impago, la tabla de efectos parciales con la definición 2 de impago se puede consultar en el anexo A.1.2.

sus componentes tienen efectos de magnitud cercana a cero o, incluso, con dirección opuesta a la esperada (concepto de interés e inflación y cálculo interés simple).

Por otra parte, los componentes del subíndice de comportamientos financieros, con excepción de la elección de productos y la elaboración de presupuesto, tienen un efecto estadísticamente significativo en la probabilidad de incumplimiento y con la dirección esperada. Entre los comportamientos, destaca que tener un balance financiero sano y el realizar compras de manera cuidadosa reducen entre 13 y 19 pp, y entre 8 y 11 pp, respectivamente, la probabilidad de caer en impago. Además, ahorrar de manera activa tiene un impacto de entre 2 y 5 pp en la probabilidad de incumplimiento, de igual forma, ponerse metas financieras a largo plazo se relaciona con una menor probabilidad de incumplimiento entre la población (Tabla 4).

Los resultados anteriores se mantienen sin cambios ante variaciones en las especificaciones del modelo, la muestra seleccionada y la definición de impago utilizada (Tabla 5 y Tabla 6). En particular, considerando controles tanto de variables socioeconómicas como de historial financiero/crediticio, las estimaciones muestran efectos significativos en la dirección esperada para los comportamientos financieros relacionado a tener ahorro activo, mantener un sano balance financiero, establecer metas financieras de largo plazo y analizar las compras que se realizan. De esta forma, la evidencia mostrada sugiere que impulsar programas que promuevan este tipo de hábitos entre la población podrían traducirse en un mejor desempeño en el mercado de crédito, lo que a su vez debería traducirse en mejores condiciones de financiamiento.

Destaca que, para ambos levantamientos de la encuesta, así como para la muestra agregada, la variable de balance financiero es la que mayor impacto tiene sobre la probabilidad de incumplimiento. De manera más puntual, un acreditado que posee un balance financiero sano tiene una probabilidad de incumplimiento entre 12 y 13 pp menor (definición uno de impago) comparada con una persona que no lo tiene (Tabla 5).

Tabla 5. Efectos de los de componentes de alfabetización financiera por período

(Definición 1 de impago)

	(2018) EDU+SE+FIN impago_1	(2021) EDU+SE+FIN impago_1	(Unión) EDU+SE+FIN impago_1
Conocimientos financieros			
Valor del dinero en el tiempo	0.0402**	-0.0371*	-0.000815
Interés e Inflación	-0.00122	-0.00259	-0.00289
Interés simple	-0.00660	0.0356*	0.0125
Interés compuesto	-0.0517**	-0.0484*	-0.0427**
Riesgo y retorno	-0.0305	-0.0222	-0.0226
Diversificación	0.0129	-0.00772	0.00126
Actitudes financieras			
Preferencia por ahorrar	-0.0163	-0.00260	-0.0107
Comportamientos financieros			
Presupuesto y poder de decisión	0.0126	-0.0181	0.000482
Ahorro activo	-0.0535**	-0.0433**	-0.0473***
Balance financiero	-0.117***	-0.133***	-0.125***
Elección de productos	-0.0215	0.00513	-0.00842
Asesoramiento independiente	-0.0326	0.0215	-0.00284
Metas a largo plazo	-0.0463**	-0.00774	-0.0276**
Compras cuidadosas	-0.0947***	-0.0560***	-0.0764***
Controles			
Socioeconómicas	Si	Si	Si
Historial financiero/crediticio	Si	Si	Si
Número de observaciones:	4,079	4,347	8,426
Población representada:	24,511,766	27,002,686	27,002,686

Nota: La variación en la probabilidad estimada se calculó para un acreditado hipotético cuyos valores en los regresores son iguales a la media poblacional para el caso de variables continuas, para variables categóricas se calculó el cambio discreto desde el nivel base. Los errores estándar se presentan entre paréntesis debajo de los efectos marginales. La significancia estadística se presenta a niveles de 10%*, del 5%** o del 1%***.

Fuente: Cálculos propios con la información de la ENIF.

Finalmente, a pesar de la robustez que muestran los resultados referentes a los comportamientos financieros, el efecto que se había encontrado en relación con el conocimiento del cálculo de interés compuesto con la definición uno de impago se pierde al cambiar por la definición dos, lo cual reafirma que son los comportamientos financieros adecuados de las personas los que podrían incidir de forma relevante en el buen desempeño de las personas acreditadas con créditos formales.

Tabla 6. Efectos de los de componentes de alfabetización financiera por período
(Definición 2 de impago)

	(2018) EDU+SE+FIN impago_2	(2021) EDU+SE+FIN impago_2	(Unión) EDU+SE+FIN impago_2
Conocimientos financieros			
Valor del dinero en el tiempo	0.0180	-0.0423**	-0.0164
Interés e Inflación	0.00588	0.00655	0.00769
Interés simple	-0.0175	0.0274*	0.00426
Interés compuesto	-0.0283	-0.0264	-0.0198
Riesgo y retorno	-0.0267	-0.0157	-0.0185
Diversificación	0.00786	0.00375	0.00523
Actitudes financieras			
Preferencia por ahorrar	-0.00480	0.0112	0.00240
Comportamientos financieros			
Presupuesto y poder de decisión	0.0195	-0.0131	0.00404
Ahorro activo	-0.0389*	-0.0199	-0.0287**
Balance financiero	-0.0749***	-0.119***	-0.0986***
Elección de productos	-0.0327**	-0.00537	-0.0209*
Asesoramiento independiente	-0.0347	0.0190	-0.000313
Metas a largo plazo	-0.0374***	-0.0125	-0.0250**
Compras cuidadosas	-0.0543***	-0.0391**	-0.0478***
Controles			
Socioeconómicas	Si	Si	Si
Historial financiero/crediticio	Si	Si	Si
Número de observaciones:	4,079	4,347	8,426
Población representada:	24,511,766	27,002,686	27,002,686

Nota: La variación en la probabilidad estimada se calculó para un acreditado hipotético cuyos valores en los regresores son iguales a la media poblacional para el caso de variables continuas, para variables categóricas se calculó el cambio discreto desde el nivel base. Los errores estándar se presentan entre paréntesis debajo de los efectos marginales. La significancia estadística se presenta a niveles de 10%*, del 5%** o del 1%***.

Fuente: Cálculos propios con la información de la ENIF.

V. Poder predictivo de la alfabetización financiera

A pesar de que los resultados parecen ser concluyentes en cuanto a la relación que existe entre los comportamientos financieros y la probabilidad de incumplimiento de los acreditados, el análisis presentado en la sección anterior no permite hacer conclusiones sobre qué tan fuerte es dicha relación. Conocer esto es fundamental para determinar si estos buenos hábitos financieros, además de que podrían mejorar las condiciones del financiamiento en el mercado de crédito, también podrían ser la clave para reducir las barreras de acceso al mismo, en particular, para la población que carece de historial financiero.

Para este análisis se realizan dos pruebas diferentes con los cuatro grupos de variables mencionados anteriormente (Figura 2): socioeconómicas, (SE); financieras y de historial crediticio, (FIN); conocimiento y actitudes (CON); y comportamientos financieros. (COM). El primero consiste en ir

suministrando cada vez más información a los modelos de puntaje crediticio para así evaluar si la alfabetización financiera de la población aporta información relevante adicional para la determinación del puntaje crediticio de las personas. La segunda prueba consiste en entrenar a los modelos con los grupos de variables por separado, esto con la finalidad de comparar el poder explicativo de las variables de alfabetización financiera respecto a las variables socioeconómicas y de historial financiero/crediticio.

Figura 2. Variables utilizadas para realizar las pruebas, segmentadas en cuatro grupos

Socioeconómicas	Historial financiero/crediticio	Conocimientos y actitudes financieras	Comportamientos financieros
<ul style="list-style-type: none"> Edad Sexo Región Situación conyugal Escolaridad Localidad Estatus laboral 	<ul style="list-style-type: none"> Cuenta de apoyo Cuenta de inversión Cantidad de cuentas de captación Rechazo de crédito Crédito informal Cantidad de créditos formales 	<ul style="list-style-type: none"> Valor del dinero en el tiempo Interés e inflación Interés simple Interés compuesto Riesgo y retorno Diversificación Preferencia por ahorrar 	<ul style="list-style-type: none"> Presupuesto Ahorro activo Balance financiero Elección de productos Asesoramiento independiente Metas a largo plazo Considera si puedo comprar

Fuente: Elaboración propia con información de la ENIF.

Para cada grupo de variables, en cada prueba, se realizan 500 simulaciones con cada uno de los modelos utilizados. En cada ocasión se entrena un modelo con una partición de muestra aleatoria para que la forma en la que se toma la muestra no influya en los resultados reportados.²⁰ Después de correr las 500 simulaciones se calcula el Área Bajo la Curva - Característica Operativa del Receptor (AUC-ROC por sus siglas en inglés) promedio.²¹ Este proceso permitirá evaluar si las variables tradicionales (socioeconómicas y de historial financiero/crediticio) tienen un mayor poder

²⁰ Se utilizan 80% de los datos para entrenamiento y 20% para prueba.

²¹ Para evaluar el rendimiento de los modelos de puntaje crediticio existen diversas métricas, las más populares son el porcentaje de observaciones correctamente clasificado, medición del error tipo I, medición del error tipo II, estadística de Kolmogorov-Smirnov, sensibilidad/recuperación, especificidad, métrico F y el área bajo la curva ROC (Dastile et al., 2020). En este trabajo se utiliza el Área Bajo la Curva - Característica Operativa del Receptor (AUC-ROC por sus siglas en inglés) como métrica para evaluar el rendimiento de los modelos, esto debido a que la literatura sugiere que es la más apropiada para tratar conjuntos de datos desbalanceados. Para calcular la métrica empleada es necesaria una matriz de confusión, la cual consiste en cuantificar los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, dependiendo de la clasificación realizada por cada modelo. Para este trabajo, los verdaderos positivos son aquellas personas que fueron etiquetadas correctamente con la etiqueta de que caerían en impago, los verdaderos negativos son aquellos individuos clasificados como que no caerían en impago correctamente, los falsos positivos son personas clasificadas como que caerían en impago pero que en realidad no caen en impago y, finalmente, los falsos negativos son aquellas personas clasificadas como que no caerían en impago pero que en realidad caen en impago. De esta forma, La métrica AUC-ROC consiste en construir una curva ROC y calcular el área bajo esta. Estas curvas muestran la relación entre la razón de verdaderos positivos y la razón de falsos positivos para todos los posibles umbrales de clasificación, dichos umbrales dictaminan el valor límite. Por lo cual, si el modelo sobrepasa este valor se clasifican como impago y si el modelo está por debajo del umbral, entonces se clasifica el individuo como no impago.

explicativo que las variables de alfabetización financiera de la población y si las nuevas metodologías de aprendizaje automatizado presentan ventajas al explotar este tipo de información para clasificar individuos en relación con la metodología clásica basada en la regresión logística.

Modelos de aprendizaje automático

Los avances en las tecnologías, el acelerado incremento en las capacidades de cómputo y la creciente cantidad de información disponible, han ocasionado que la diversidad de variables y metodologías para el análisis del puntaje crediticio se incremente significativamente. Lo anterior se ha beneficiado por el espíritu innovador de los nuevos participantes que llegan al mercado financiero y ha contribuido a hacer más eficiente la gestión del riesgo de crédito en la operación de intermediarios financieros, al facilitar la obtención y procesamiento de información de clientes. Lo cual ha beneficiado el acceso al crédito, al automatizar los procesos de aprobación, gracias a que logran clasificar de forma cada vez más eficiente a los clientes potenciales (Bartlett et al., 2019; Berg et al., 2020). No obstante, la sofisticación en la medición del riesgo de crédito también ha representado retos cada vez mayores para las autoridades supervisoras, quienes tienen que estar en constante actualización de su marco regulatorio para mantenerse a la vanguardia, pero sin perder de vista que las metodologías empleadas ofrezcan mayor eficiencia y transparencia para su supervisión (Alonso & Carbó, 2020).

El aprendizaje automático o aprendizaje máquina (del inglés, *machine learning*) se basa en métodos computacionales que realizan predicciones basadas en “experiencias” y algoritmos estadísticos. En este ámbito, la experiencia se refiere a datos con información pasada que alimentan el algoritmo, los cuales se analizan para obtener información acerca de tendencias, comportamientos y otras relaciones que pueden ser utilizadas para la predicción (Mohri et al., 2012).

Al conjunto de datos con la información pasada se le denomina conjunto de entrenamiento. Dentro del campo del aprendizaje automático existen diversas formas de utilizar esta experiencia para extraer la información necesaria que ayude a la predicción. A cada una de estas formas se le conoce como “tipo de aprendizaje”, en donde los más comunes son los siguientes:

- **Aprendizaje supervisado:** los datos de entrenamiento consisten en pares de vectores, donde uno de ellos es el vector de entrada que contiene la información con las características que serán utilizadas para realizar la predicción. El otro vector contiene la información de los resultados deseados, es decir, los modelos de aprendizaje automático que trabajan bajo esta filosofía tienen el objetivo de crear una función

capaz de predecir el resultado correspondiente a cualquier vector de entrada después de haber visto una serie de ejemplos (los datos de entrenamiento).

- **Aprendizaje no supervisado:** En este grupo de modelos, para cada entrada en el conjunto de datos de entrenamiento no se brinda ninguna salida, por lo que se le deja al modelo la tarea de encontrar algún tipo de patrón que ayude a agrupar los valores de entrada.

Los modelos de puntaje crediticio basados en algoritmos de aprendizaje automático que se utilizan en este trabajo son de aprendizaje supervisado, ya que es posible extraer un conjunto de datos de entrenamiento ya etiquetados con la salida correspondiente. Las metodologías de aprendizaje automático son diversas tanto en antigüedad como en complejidad y, dado que no es el objetivo de este trabajo ser exhaustivo en la explicación de las distintas metodologías existentes, basta con señalar que en esta sección se utilizan cinco métodos diferentes, se utiliza la regresión logística y otros cuatro métodos de aprendizaje automático que son calibrados con el proceso señalado anteriormente para la obtención de sus parámetros: bosques aleatorios, AdaBoost, redes neuronales y k-vecinos más cercanos.²²

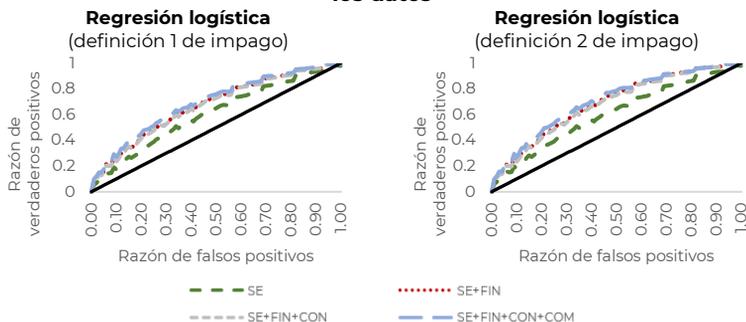
²² Aunque no se pretende ser exhaustivo en esta revisión, aquí se presenta una breve explicación de las metodologías y los parámetros utilizados en este trabajo:

- Los **bosques aleatorios** *bagging* pertenecen a la categoría de modelos de ensamblaje. La finalidad de este tipo de modelos es combinar varios clasificadores de un tipo que resuelvan la tarea que se desea realizar (en este caso, clasificar en una variable binaria a los individuos para su probabilidad de incumplimiento), para así poder incrementar el rendimiento de la clasificación (Breiman, 1996). Para este trabajo se utilizan 101 árboles de decisión como clasificadores, por lo que, se emplea la técnica de *bootstrap* para crear 101 muestras distintas y con cada muestra entrenar los 101 árboles de decisión, para clasificar a los individuos. Al final, se etiqueta a un individuo dependiendo de lo que la mayoría de los 101 clasificadores hayan predicho.
- El *boosting*, desarrollado por Schapire, (1990), también pertenece a la categoría de métodos de ensamblaje. El modelo más popular de este tipo se llama **Adaptive Boosting, mejor conocido como AdaBoost**, el cual fue desarrollado por Freund & Schapire, (1997). En este trabajo se utilizan árboles de decisión como clasificadores a "ensamblar", en total se utilizan 50 árboles de decisión, los cuales se retroalimentan entre sí. Cada que se entrena un árbol se le da más peso a aquellos individuos que el árbol anterior haya etiquetado mal y así sucesivamente. Al final, se hace una votación donde participan los 50 árboles para asignarle una clase (impago o no impago) al individuo evaluado, dándole más importancia a los votos de aquellos árboles que tengan el peso más grande asociado a dicho individuo.
- Los modelos de **redes neuronales** artificiales se han popularizado desde que Werbos, (1974) desarrollara un método para su entrenamiento. Este método se conoce como *Back-Propagation* y consiste en calcular las derivadas parciales de los pesos de una red neuronal para que, con base en esa información, se logre un ajuste de pesos que minimice la función de pérdida. Una red neuronal artificial consiste en un conjunto de nodos que están conectados entre sí, estos nodos son llamados neuronas y la forma en la que se conectan trata de simular a las conexiones de una red neuronal biológica. La red neuronal se organiza por capas, donde cada capa tiene una cierta cantidad de neuronas y cada neurona dentro de la capa está enlazada a otras neuronas que se encuentran en la siguiente capa, en dichos enlaces el valor de salida de cada neurona es multiplicado por un valor numérico denominado "peso" el cual determina el nivel de importancia que la siguiente neurona tiene que dar a esa información, a la combinación lineal de los pesos y valores de entrada se le suma un valor extra denominado "sesgo". De igual manera, en la salida de cada neurona existe una función que modifica el resultado de esta antes de que dicho valor sea propagado a la red, esta función es llamada función de activación. Después de que la información sea propagada por toda la red, esta arroja una predicción, la cual es comparada con el valor real

Grupos agregados

Para todas las especificaciones y modelos se realizaron las pruebas con ambas definiciones de impago. Las métricas ROC-AUC obtenidas muestran que las variables de historial crediticio agregan información relevante a las variables socioeconómicas para mejorar la precisión en la clasificación de individuos que potencialmente podrían caer en el impago de sus obligaciones financieras. Además, los comportamientos financieros parecen aportar información nueva a estos dos grupos de variables, que permite diferenciar entre un acreditado que cae en impago y uno que no. El grupo de variables de conocimientos financieros no muestra evidencia de aportar información adicional a la que ya suministran los grupos de variables comúnmente utilizados, ya que, las curvas ROC asociadas a los modelos entrenados con los grupos de variables SE+FIN y SE+FIN+CON parecen encimarse (Figura 3).

Figura 3. Curvas ROC promedio de grupos agregados con la Unión de los datos



Fuente: Cálculos propios con información de la ENIF.

Esto puede verificarse también con el área bajo la curva de ambos grupos, ya que esta no muestra diferencias significativas. La diferencia

mediando el error mediante una función de pérdida. El modelo se basa en minimizar la función de pérdida modificando los pesos y sesgos de la red, a partir del algoritmo de *Back-Propagation*. En este trabajo se utiliza una red neuronal de una sola capa con cinco neuronas en ella, la función de activación que tiene cada neurona es la función tangente hiperbólica (*tanh*), se entrena durante 400 iteraciones y se utiliza el optimizador Adam para ajustar los pesos de la red. Cabe mencionar que las especificaciones del modelo antes mencionadas se establecieron de dicha manera por ser la combinación que mejor desempeño tuvo en el período de entrenamiento.

- El **modelo de k-vecinos** más cercanos fue desarrollado por Fix & Hodges, (1989) y consiste en que dado un individuo con un vector de características x , el algoritmo calcula la distancia a los k puntos (individuos) más cercanos y le asigna la clasificación de impago o no impago dependiendo de la etiqueta que tengan la mayoría de los individuos más cercanos a dicho punto, cada punto vota por la clase a la que pertenece, y la clase con más votos se toma como la predicción. En este trabajo se calcula la distancia a los 51 "vecinos más cercanos", y se utiliza como medida de distancia la de Minkowski (se utiliza esta medida por ser la que mejores resultados mostró en la etapa de entrenamiento) definida como:

$$d = \left[\sum_{i=1}^n (|x_i - x'_i|)^k \right]^{\frac{1}{k}},$$

donde n es el número de características del individuo evaluado, x_i es la característica i –ésima del individuo evaluado x , y x'_i es la característica i –ésima del individuo x' (perteneciente al conjunto de entrenamiento).

máxima entre dichas áreas es de 0.12 puntos porcentuales (AUC promedio con definición dos de impago). Los resultados son robustos ante ambas definiciones de impago. Los resultados también muestran que los modelos de puntaje crediticio son más precisos cuando se agrega el grupo de variables de comportamientos financieros, aunque la curva ROC asociada a este grupo (SE+FIN+CON+COM) presenta solo incrementos marginales respecto a las dos curvas ROC que la preceden (Tabla 7).

Tabla 7. Áreas bajo las curvas ROC promedio con grupos agregados, por modelo, definición de impago y grupo de variables

Modelo	Grupo de variables	Definición 1 de impago		Definición 2 de impago	
		AUC promedio	Desviación estándar	AUC promedio	Desviación estándar
Regresión logística	SE	60.31%	1.94%	60.94%	2.31%
	SE+FIN	67.36%	2.01%	69.55%	1.87%
	SE+FIN+CON	67.65%	2.05%	69.52%	2.01%
	SE+FIN+CON+COM	70.65%	1.96%	72.56%	1.91%
Red neuronal	SE	60.41%	2.03%	60.46%	2.35%
	SE+FIN	67.05%	2.08%	68.79%	2.10%
	SE+FIN+CON	66.60%	1.98%	67.95%	2.21%
	SE+FIN+CON+COM	69.16%	1.89%	70.63%	2.08%
AdaBoost	SE	60.07%	1.85%	61.12%	2.20%
	SE+FIN	67.75%	1.99%	69.66%	2.12%
	SE+FIN+CON	67.88%	2.00%	69.81%	2.11%
	SE+FIN+CON+COM	70.89%	1.84%	72.50%	2.10%
K-Vecinos	SE	59.22%	1.92%	59.82%	2.38%
	SE+FIN	65.61%	1.86%	67.01%	2.01%
	SE+FIN+CON	65.65%	1.98%	66.09%	1.91%
	SE+FIN+CON+COM	68.53%	1.83%	68.96%	2.02%
Bosque aleatorio	SE	55.37%	1.99%	55.88%	2.45%
	SE+FIN	62.88%	1.78%	65.68%	2.01%
	SE+FIN+CON	64.68%	1.99%	66.54%	1.94%
	SE+FIN+CON+COM	69.10%	1.91%	70.71%	1.84%

Fuente: Cálculos propios con información de la ENIF.

Para comprobarlo se realizaron pruebas de diferencia de medias de cola derecha entre cada grupo y el resto. Se toma la diferencia del AUC poblacional del grupo de la izquierda respecto al AUC poblacional del grupo de la parte superior. Así, si comparamos el modelo entrenado con el grupo de variables SE+FIN+CON+COM (definición uno de impago) contra el modelo entrenado con el grupo SE+FIN, se concluye que la eficiencia del modelo es mayor cuando se le suministra la información de las variables SE+FIN+CON+COM en comparación a si solo se le suministra la información de las variables SE+FIN (Tabla 8). Bajo este mismo razonamiento, para ambas definiciones de impago, agregar las variables de conocimientos financieros no parece aportar información adicional al modelo entrenado con las variables SE+FIN.

Tabla 8. Valores Z de pruebas de hipótesis de cola derecha para áreas bajo las curvas ROC promedio con grupos agregados
(Definición 1 de impago)

Modelo	Grupo de variables	CONTROL	SE	SE+FIN	SE+FIN+CON
Regresión logística	SE	118.55***			
	SE+FIN	192.96***	56.36***		
	SE+FIN+CON	192.87***	58.17***	2.26**	
	SE+FIN+CON+COM	235.65***	83.82***	26.25***	23.72***
Red neuronal	SE	114.72***			
	SE+FIN	183.25***	51.12***		
	SE+FIN+CON	187.75***	48.87***	-3.53	
	SE+FIN+CON+COM	227.18***	70.65***	16.78***	20.95***
AdaBoost	SE	122.08***			
	SE+FIN	199.1***	63.2***		
	SE+FIN+CON	200.17***	64.17***	0.99	
	SE+FIN+CON+COM	254.54***	92.96***	25.93***	24.87***
K-Vecinos	SE	107.16***			
	SE+FIN	187.7***	53.37***		
	SE+FIN+CON	176.46***	51.99***	0.30	
	SE+FIN+CON+COM	226.44***	78.41***	25.07***	23.94***
Bosque aleatorio	SE	60.24***			
	SE+FIN	161.57***	62.79***		
	SE+FIN+CON	164.77***	73.85***	15.05***	
	SE+FIN+CON+COM	223.05***	111.03***	53.12***	35.73***

Nota: Se rechaza la hipótesis nula al con los niveles de significancia de 10%*, del 5%** o del 1%***.

Fuente: Cálculos propios con la información de la ENIF.

Estos resultados confirman que las variables de comportamientos financieros aportan nueva información relevante para poder diferenciar entre una persona que potencialmente incumple en el pago de sus créditos de una persona que no lo hace, mientras que las variables de conocimientos financieros no parecen aportar información adicional. Por otra parte, no hay evidencia significativa de que los modelos de puntaje crediticio de aprendizaje automático tengan un mejor desempeño que la regresión logística, en términos de eficiencia del modelo. Sin embargo, esto está sujeto a la muestra considerada y deja abierta la pregunta si con grandes volúmenes de información podría observarse un mejor desempeño para ese tipo de metodologías.

A pesar de lo anterior, la amplia variedad de modelos de puntaje crediticio utilizados permite comprobar la robustez de los resultados. Ante ambas definiciones de impago, los resultados obtenidos con los modelos de puntaje crediticio basados en algoritmos de aprendizaje automático son consistentes con los encontrados con el modelo de regresión logística. Todos los modelos de puntaje crediticio arrojan que las variables de comportamientos financieros aportan información relevante adicional a la ya extraída de las variables de uso común (socioeconómicas y de historial financiero/crediticio). Además, los conocimientos financieros no contribuyen de forma estadísticamente significativa a incrementar la precisión de las predicciones de los modelos (Anexo 2.1).

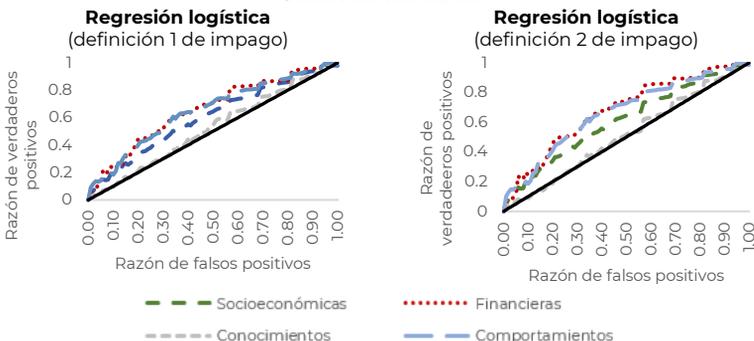
Tomando como punto de partida los modelos entrenados con la información empleada usualmente (SE+FIN), los resultados sugieren que agregar al análisis crediticio de las personas la información sobre sus comportamientos financieros mejora la precisión de los modelos de puntaje crediticio, mientras que la información sobre el conocimiento financiero no tiene efectos estadísticamente significativos sobre la predicción de los modelos. Solo en el modelo de puntaje crediticio basado en el algoritmo de Bosque aleatorio se observa una mejora en la precisión al momento de agregar información sobre los conocimientos financieros del individuo.

Grupos independientes

Aunque se ha dado evidencia que sugiere la contribución que tiene la información referente a los comportamientos financieros de la población para la medición de su probabilidad de incumplimiento en el pago de sus obligaciones financieras, el análisis realizado hasta el momento no permite concluir si algún grupo de información tiene un mayor poder predictivo para el análisis del puntaje crediticio. En aras de responder a esto, se realizan las pruebas con los grupos de variables de manera independiente.

En congruencia con la sección previa, para ambas definiciones de impago, los conocimientos financieros son el grupo de variables que menos poder explicativo poseen para poder discernir entre un individuo que cae en impago y uno que no. No obstante, los comportamientos financieros muestran un poder explicativo similar al grupo de información financiera/crediticia. Y ambos grupos son los que mayor precisión otorgan en la predicción de los modelos de puntaje crediticio, tanto de aprendizaje automático como para la regresión logística. Para ambas definiciones de impago, las variables que más contribuyen con la eficiencia de la predicción de los modelos son las de historial financiero/crediticio, seguidas de la información de los comportamientos financieros (Figura 4).

Figura 4. Curvas ROC promedio de grupos independientes con la Unión de los datos



Fuente: Cálculos propios con información de la ENIF.

Los resultados de las curvas ROC sugieren que las variables de comportamientos financieros tienen un alto poder explicativo de la probabilidad de incumplimiento de pago de las personas, el cual se asemeja a las variables de historial financiero/crediticio. Además, los resultados son robustos ante cambios en la definición de impago y en la metodología utilizada. Con todas las metodologías, las variables de comportamientos financieros muestran mayor poder predictivo que las variables referentes a los conocimientos financieros y respecto a las variables socioeconómicas. Destaca que la información sobre los conocimientos financieros resulta apenas mejor que lanzar una moneda al aire para decidir si el acreditado cae en impago o no. Sin embargo, esto último podría deberse a que ese tipo de variables podrían haber contribuido a la decisión de acceder al mercado crediticio en una primera instancia, sin embargo, por la naturaleza de la muestra utilizada no es posible determinarlo.

Tabla 9. Valores Z de pruebas de hipótesis de cola derecha para áreas bajo las curvas ROC promedio con grupos independientes
(Definición 1 de impago)

Modelo	Grupo de variables	CONTROL	SE	FIN	CON
Regresión logística	SE	112.2***			
	FIN	181.4***	47.05***		
	CON	24.36***	-58.30	-104.10	
	COM	180.2***	38.80***	-9.98	97.92***
Red neuronal	SE	116.3***			
	FIN	180.8***	45.91***		
	CON	19.80***	-64.20	-108.10	
	COM	168.1***	39.58***	-5.70	101.3***
AdaBoost	SE	114.8***			
	FIN	182.5***	47.42***		
	CON	28.05***	-60.80	-108.00	
	COM	187.2***	38.67***	-11.30	102.6***
K-Vecinos	SE	111.4***			
	FIN	151.7***	36.42***		
	CON	9.473***	-63.80	-94.60	
	COM	148.4***	29.90***	-7.08	89.84***
Bosque aleatorio	SE	59.67***			
	FIN	165.8***	68.37***		
	CON	23.54***	-26.10	-96.40	
	COM	153.4***	61.36***	-6.71	89.12***

Nota: Se rechaza la hipótesis nula al con los niveles de significancia de 10%*, del 5%** o del 1%***.

Fuente: Cálculos propios con la información de la ENIF.

Finalmente, al igual que en el ejercicio de grupos agregados, aunque ante alguna de las definiciones de impago algunas metodologías de aprendizaje automático presentan mayor precisión que la regresión logística clásica (*Adaboost* y *Red Neuronal*), los resultados no son concluyentes. Por lo que, dada la base de datos utilizada, los modelos de aprendizaje automático parecen no incrementar de forma significativa la eficiencia para predecir si un individuo caerá en impago (Anexo 2.2).

VI. Consideraciones finales

El análisis del riesgo de crédito es un tema de investigación abierto. A pesar de la amplia literatura existente, se ha dinamizado recientemente por el surgimiento de nuevas metodologías, la mayor disponibilidad de información y el estudio de nuevos componentes que pudieran incidir en el desempeño financiero de la población. La constante expansión del estudio de las causas y retos de la inclusión financiera han abierto un nuevo panorama hacia variables que podrían incidir en el desempeño financiero de la población. En este escenario, la educación financiera ha comenzado a jugar un papel cada vez más relevante, debido a sus efectos positivos sobre los conocimientos, los comportamientos y las actitudes financieras de la población.

Los resultados de este trabajo sugieren que los comportamientos financieros de la población pueden incidir de forma significativa en su buen desempeño en el mercado de crédito, aunque se encuentra un efecto no significativo de los conocimientos financieros. Esto podría aportar información valiosa para los encargados del desarrollo de los programas de educación financiera que buscan promover el acceso al mercado crediticio. En particular, fomentar el desarrollo de buenos comportamientos financieros, como el establecimiento de metas financieras de largo plazo, procurar un balance adecuado entre ingresos y gastos, y desarrollar el hábito del ahorro, se reflejaría en un menor riesgo de crédito percibido por las instituciones de crédito y, a su vez, en mejores condiciones y menores costos de financiamiento.

Además, la evidencia sugiere que los comportamientos financieros pueden dar una buena señal del desempeño futuro de los acreditados y ofrecen un poder de predicción solo menor al observado en las variables de historial financiero y crediticio. Sin duda, estos resultados cobran mayor relevancia cuando se trata de personas no incluidas en el sistema financiero, ya que, la información de los niveles de alfabetización financiera de los solicitantes de crédito, en particular la de sus comportamientos, podría ser la clave para que esta población pueda acceder al financiamiento, lo cual, incentivaría el incremento de la penetración de los servicios financieros en el país y en el bienestar económico de la población.

Referencias

- Abdou, H., & Pointon, J. (2011). Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature. *Int. Syst. in Accounting, Finance and Management*, 18, 59-88. <https://doi.org/10.1002/isaf.325>
- Alonso, A., & Carbó, J. M. (2020). Machine Learning in Credit Risk: Measuring the Dilemma Between Prediction and Supervisory Cost. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3724374>
- Assef, F. M., & Steiner, M. T. A. (2020). Ten-year evolution on credit risk research: A systematic literature review approach and discussion. *Ingeniería e Investigación*, 40(2), 50-71. <https://doi.org/10.15446/ing.investig.v40n2.78649>
- Babaev, D., Savchenko, M., Tuzhilin, A., & Umerenkov, D. (2019). E.T.-RNN: Applying Deep Learning to Credit Loan Applications. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2183-2190. <https://doi.org/10.1145/3292500.3330693>
- Banerjee, A., Duflo, E., Glennerster, R., & Kinnan, C. (2015). The miracle of microfinance? Evidence from the randomized evaluation. *American Economic Journal: Applied Economics*, 7(1), 22-53.
- Bank for International Settlements. (2005). An Explanatory Note on the Basel II IRB Risk Weight Functions. Basel Committee on Banking Supervision.
- Bank for International Settlements. (2006). International Convergence of Capital Measurement: A Revised Framework. Basel Committee on Banking Supervision.
- Barajas, A., Beck, T., Belhaj, M., & Ben Naceur, S. (2020). Financial inclusion: What have we learn so far? What do we have to learn? <https://www.imf.org/en/Publications/WP/Issues/2020/08/07/Financial-Inclusion-What-Have-We-Learned-So-Far-What-Do-We-Have-to-Learn-49660>
- Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). Consumer-Lending Discrimination in the FinTech Era (Working Paper N.º 25943). National Bureau of Economic Research. <https://doi.org/10.3386/w25943>
- Beck, T., Demirguc-Kunt, A., & Honohan, P. (2009). Access to financial services: Measurement, impact, and policies. *World Bank Research Observer*, 24(1), 119-145.
- Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the Rise of FinTechs: Credit Scoring Using Digital Footprints. *The Review of Financial Studies*, 33(7), 2845-2897. <https://doi.org/10.1093/rfs/hhz099>
- Boyes, W. J., Hoffman, D. L., & Low, S. A. (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics*, 40(1), 3-14. [https://doi.org/10.1016/0304-4076\(89\)90026-2](https://doi.org/10.1016/0304-4076(89)90026-2)
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140. <https://doi.org/10.1007/BF00058655>



- Breiman, L. (2001). Random Forests.
- Brown, M., Grigsby, J., van der Klaauw, W., Wen, J., & Zafar, B. (2016). Financial Education and the Debt Behavior of the Young. *The Review of Financial Studies*, 29(9), 2490-2522. <https://doi.org/10.1093/rfs/hhw006>
- Bruhn, M., Lara Ibarra, G., & McKenzie, D. (2014). The minimal impact of a large-scale financial education program in Mexico City. *Journal of Development Economics*, 108, 184-189. <https://doi.org/10.1016/j.jdeveco.2014.02.009>
- Bruhn, M., Leão, L. de S., Legovini, A., Marchetti, R., & Zia, B. (2016). The Impact of High School Financial Education: Evidence from a Large-Scale Evaluation in Brazil. *American Economic Journal: Applied Economics*, 8(4), 256-295. <https://doi.org/10.1257/app.20150149>
- Chen, H., & Xiang, Y. (2017). The Study of Credit Scoring Model Based on Group Lasso. *Procedia Computer Science*, 122, 677-684. <https://doi.org/10.1016/j.procs.2017.11.423>
- Chorafas, D. (2000). *Managing credit risk, analysing rating and pricing the probability of default*. Londres: Euromoney Institutional Investor PLC.
- Comisión Nacional Bancaria y de Valores. (s. f.). Disposiciones de carácter general aplicables a las instituciones de crédito. Recuperado 8 de septiembre de 2022, de <https://www.cnbv.gob.mx/Normatividad/Disposiciones%20de%20car%C3%A1cter%20general%20aplicables%20a%20las%20instituciones%20de%20cr%C3%A9dito.pdf>
- Comisión Nacional Bancaria y de Valores. (2019). *Alfabetización Financiera en México*.
- Comisión Nacional Bancaria y de Valores (2020). *¿Cuáles son los impactos de la inclusión financiera en las personas y las empresas?*.
- Comisión Nacional Bancaria y de Valores. (2021). *El crédito en México: Productos, instrumentos y evolución* (p. 58).
- R04 Cartera de crédito. R04 H créditos a la vivienda. Instructivo de llenado, (2021).
- Comisión Nacional Bancaria y de Valores. (2022). *Encuesta Nacional de Inclusión Financiera (ENIF) 2021: Reporte de resultados*.
- Comité de Educación Financiera. (2017). *Estrategia Nacional de Educación Financiera*.
- Cull, R., Ehrbeck, T., & Holle, N. (2014). Financial inclusion and development: Recent impact evidence. CGAP. <https://www.cgap.org/sites/default/files/FocusNote-Financial-Inclusion-and-Development-April-2014.pdf>
- Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, 91, 106263. <https://doi.org/10.1016/j.asoc.2020.106263>
- Demirguc-Kunt, A., Klapper, L., & Singer, D. (2017). Financial inclusion and inclusive growth: A review of empirical evidence. <https://documents1.worldbank.org/curated/en/403611493134249446/pdf/WPS8040.pdf>

- Farhat, J. B., & Robb, A. (2014). Applied Survey Data Analysis Using Stata: The Kauffman Firm Survey Data. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.2477217>
- Fix, E., & Hodges, J. L. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review / Revue Internationale de Statistique*, 57(3), 238-247. <https://doi.org/10.2307/1403797>
- Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. <https://doi.org/10.1006/jcss.1997.1504>
- Fujiki, H. (2020). Cash demand and financial literacy: A case study using Japanese survey data. *Japan and the World Economy*, 54, 100998. <https://doi.org/10.1016/j.japwor.2020.100998>
- García, N., Grifoni, A., López, J. C., & Mejía, D. (2013). N° 12. La educación financiera en América Latina y el Caribe. Situación actual y perspectivas. CAF. <https://cafsciotea.azurewebsites.net/handle/123456789/379>
- Glennon, D., Kiefer, N., Larson, C. E., & Choi, H. (2008). Development and validation of credit scoring models. *The Journal of Credit Risk*, 4(3), 41-101. <https://doi.org/10.21314/JCR.2008.075>
- Greene, W. H. (1992). A Statistical Model for Credit Scoring (SSRN Scholarly Paper N.º 1867088). <https://papers.ssrn.com/abstract=1867088>
- Gutierrez Girault, M. A. (2007, octubre). Modelos de credit scoring: Qué, cómo, cuándo y para qué [MPRA Paper]. <https://mpra.ub.uni-muenchen.de/16377/>
- Heeringa, S., West, B. T., & Berglund, P. A. (2017). Applied survey data analysis (Second edition). CRC Press, Taylor & Francis Group.
- Huang, C.-L., Chen, M.-C., & Wang, C.-J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications*, 33(4), 847-856. <https://doi.org/10.1016/j.eswa.2006.07.007>
- Hvidberg, K. B. (2022). Field of Study and Financial Problems: How Economics Reduces the Risk of Default.
- Kaiser, T., Lusardi, A., Menkhoff, L., & Urban, C. (2022). Financial education affects financial knowledge and downstream behaviors. *Journal of Financial Economics*, 145(2, Part A), 255-272. <https://doi.org/10.1016/j.jfineco.2021.09.022>
- Kaiser, T., & Menkhoff, L. (2017). Does financial education impact financial literacy and financial behavior, and if so, when?
- Kaiser, T., & Menkhoff, L. (2020). Financial education in schools: A meta-analysis of experimental studies. *Economics of Education Review*, 78, 101930. <https://doi.org/10.1016/j.econedurev.2019.101930>
- Loukoianova, E., & Yang, Y. (2018). Financial Inclusion in Asia-Pacific. No. 18/17. <https://www.imf.org/en/Publications/Departmental-Papers-Policy-Papers/Issues/2018/09/18/Financial-Inclusion-in-Asia-Pacific-46115>



- Louzada, F., Ara, A., & Fernandes, G. B. (2016). Classification methods applied to credit scoring: A systematic review and overall comparison (arXiv:1602.02137). arXiv. <http://arxiv.org/abs/1602.02137>
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). Foundations of machine learning. MIT Press.
- Mungaray, A., Gonzalez Arzabal, N., & Osorio Novela, G. (2021). Educación financiera y su efecto en el ingreso en México. *Problemas del Desarrollo. Revista Latinoamericana de Economía*, 52(205). <https://doi.org/10.22201/iiec.20078951e.2021.205.69709>
- Organisation for Economic Co-operation and Development. (2005). Improving Financial Literacy: Analysis of Issues and Policies. OECD. https://www.oecd-ilibrary.org/finance-and-investment/improving-financial-literacy_9789264012578-en
- Organisation for Economic Co-operation and Development. (2013). OECD Glossary of Statistical Terms—Credit Definition. <https://stats.oecd.org/glossary/detail.asp?ID=474>
- Organisation for Economic Co-operation and Development. (2017). G20/OECD INFE report on adult financial literacy in G20 countries. <https://www.oecd.org/finance/g20-oecd-infe-report-adult-financial-literacy-in-g20-countries.htm>
- Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Klamargias, A. (2018). A robust machine learning approach for credit risk analysis of large loan-level datasets using deep learning and extreme gradient boosting. 45.
- Pitblado, J. (2009). Survey Data Analysis in Stata. 28.
- Sahay, M., & Chiak, R. (2020). Finance and Inequality. IMF. Staff Discussion Note SDN/20/01. <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2020/01/16/Finance-and-Inequality-45129>
- Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, 5(2), 197-227. <https://doi.org/10.1007/BF00116037>
- Urban, C., Schmeiser, M., Collins, J. M., & Brown, A. (2020). The effects of high school personal financial education policies on financial behavior. *Economics of Education Review*, 78, 101786. <https://doi.org/10.1016/j.econedurev.2018.03.006>
- Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences.

Anexo 1. Efectos marginales

Tabla A.1.1. Efectos de la alfabetización financiera

(Definición 2 de impago)

	(1) EDU <i>impago_2</i>	(2) EDU+SE <i>impago_2</i>	(3) EDU+FIN <i>impago_2</i>	(4) EDU+SE+FIN <i>impago_2</i>
Alfabetización financiera				
Conocimientos financieros	-0.00917	-0.00784	-0.0104	-0.00753
Actitudes financieras	-0.000650	-0.000297	0.000177	0.000607
Comportamientos financieros	-0.0620***	-0.0583***	-0.0493***	-0.0464***
Socioeconómicas				
Edad de 30 a 44		-0.00726		-0.00604
Edad de 45 a 59		-0.0461**		-0.0350**
Edad de 60 o más		-0.123***		-0.0860***
Sexo Mujer		-0.0182		-0.0140
Localidad Urbana		0.0568***		0.0353***
Escolaridad Secundaria		0.0323*		0.0224
Escolaridad Media superior		0.0256		0.00564
Escolaridad Licenciatura o más		0.0209		-0.00580
Situación conyugal Separado		0.0621***		0.0444**
Situación conyugal Soltero		-0.0339**		-0.0202
Región Centro Sur y Oriente		-0.0559***		-0.0522***
Región Ciudad de México		-0.0737***		-0.0453**
Región Occidente y Bajío		-0.0726***		-0.0504***
Región Noreste		-0.0349*		-0.0261*
Región Noroeste		-0.0151		-0.0145
Estatus lab. Trabajador informal		0.0453**		0.0113
Estatus lab. Trabajador formal		-0.0287*		-0.0466***
Historial financiero/crediticio				
Tiene cuenta de apoyo			-0.0177	-0.00687
Tiene cuenta de inversión			-0.00844	-0.0156
Núm. de cuentas de captación			-0.0119	-0.00311
Rechazo anterior de crédito			0.132***	0.120***
Tiene crédito informal			0.0965***	0.0853***
Núm. de créditos formales			0.0238***	0.0254***
<i>Número de observaciones:</i>	8,426	8,426	8,426	8,426
<i>Población representada:</i>	27,002,686	27,002,686	27,002,686	27,002,686

Nota: La variación en la probabilidad estimada se calculó para un acreditado hipotético cuyos valores en los regresores son iguales a la media poblacional para el caso de variables continuas, para variables categóricas se calculó el cambio discreto desde el nivel base. Los errores estándar se presentan entre paréntesis debajo de los efectos marginales. La significancia estadística se presenta a niveles de 10%*, del 5%** o del 1%***.

Fuente: Cálculos propios con la información de la ENIF.

Tabla A.1.2. Efectos de los componentes de la alfabetización financiera
(Definición 2 de impago)

	(1) EDU impago_2	(2) EDU+SE impago_2	(3) EDU+FIN impago_2	(4) EDU+SE+FIN impago_2
Conocimientos financieros				
Valor del dinero en el tiempo	-0.0247	-0.0181	-0.0226	-0.0164
Interés e Inflación	0.0163	0.0166	0.00525	0.00769
Interés simple	0.00221	-0.000206	0.00391	0.00426
Interés compuesto	-0.0220	-0.0135	-0.0266*	-0.0198
Riesgo y retorno	-0.0231	-0.0232	-0.0187	-0.0185
Diversificación	0.00520	0.00594	0.00521	0.00523
Actitudes financieras				
Preferencia por ahorrar	-0.000188	0.00221	0.00136	0.00240
Comportamientos financieros				
Presupuesto y poder de decisión	0.00264	0.00657	-0.00220	0.00404
Ahorro activo	-0.0125	-0.0176	-0.0262**	-0.0287**
Balance financiero	-0.168***	-0.157***	-0.102***	-0.0986***
Elección de productos	-0.0139	-0.0100	-0.0235**	-0.0209*
Asesoramiento independiente	-0.00223	0.000712	-0.00258	-0.000313
Metas a largo plazo	-0.0252**	-0.0288**	-0.0227**	-0.0250**
Compras cuidadosas	-0.0752***	-0.0678***	-0.0516***	-0.0478***
Controles				
Socioeconómicas	No	Si	No	Si
Historial financiero/crediticio	No	No	Si	Si
Número de observaciones:	8,426	8,426	8,426	8,426
Población representada:	27,002,686	27,002,686	27,002,686	27,002,686

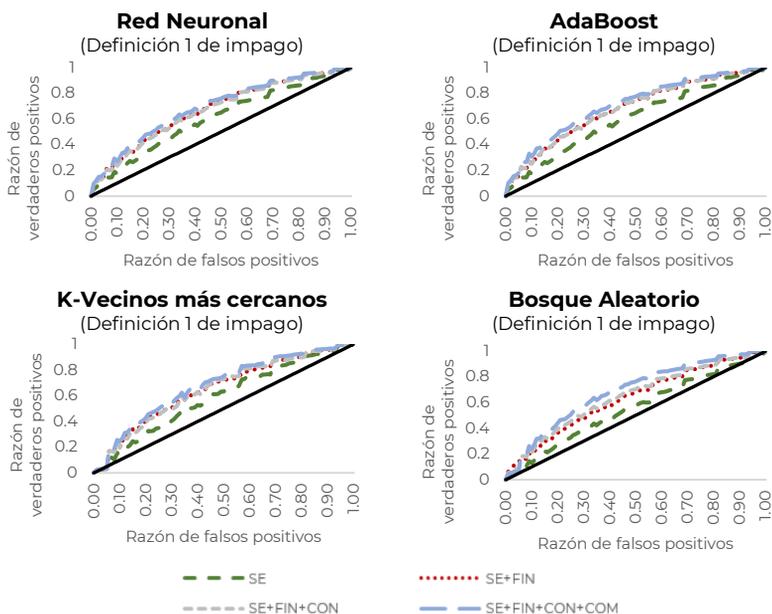
Nota: La variación en la probabilidad estimada se calculó para un acreditado hipotético cuyos valores en los regresores son iguales a la media poblacional para el caso de variables continuas, para variables categóricas se calculó el cambio discreto desde el nivel base. Los errores estándar se presentan entre paréntesis debajo de los efectos marginales. La significancia estadística se presenta a niveles de 10%*, del 5%** o del 1%***.

Fuente: Cálculos propios con la información de la ENIF.

Anexo 2. Curvas ROC promedio con la Unión de los datos

Anexo 2.1. Grupos agregados

Los resultados mostrados en este anexo son el promedio de 500 simulaciones generadas de forma independiente y aleatoria empleando la muestra conjunta de las ENIF 2018 y 2021. El 80% de los datos se utilizaron para entrenar los modelos y el 20% restante se utilizaron para realizar las pruebas de rendimiento en cada iteración. Se realizó el mismo ejercicio para cada uno de los levantamientos por separado y los resultados muestran el mismo comportamiento.²³

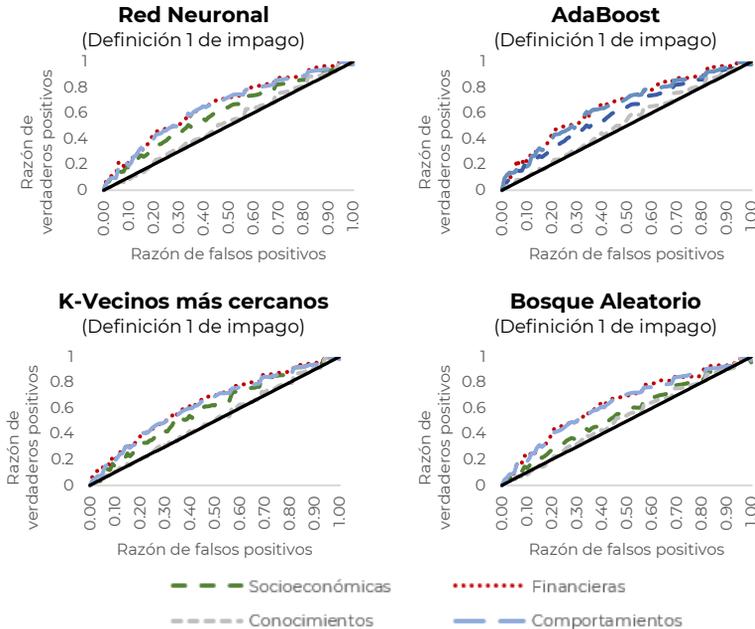


Fuente: Cálculos propios con información de la ENIF.

²³ Los gráficos y estadísticos de dichas simulaciones están disponibles a petición directa a los autores.

Anexo 2.2. Grupos independientes

Los resultados mostrados en este anexo son el promedio de 500 simulaciones generadas de forma independiente y aleatoria empleando la muestra conjunta de las ENIF 2018 y 2021. El 80% de los datos se utilizaron para entrenar los modelos y el 20% restante se utilizaron para realizar las pruebas de rendimiento en cada iteración. Se realizó el mismo ejercicio para cada uno de los levantamientos por separado y los resultados muestran el mismo comportamiento.²⁴



Fuente: Cálculos propios con información de la ENIF.

²⁴ Los gráficos y estadísticos de dichas simulaciones están disponibles a petición directa a los autores.