



Revista Venezolana de Gerencia





Crédito hipotecario: un modelo predictivo de discriminación de riesgo

González-Rossano, Carlos*
De la Torre-Díaz, Lorena**
Terán-Bustamante, Antonia***

Resumen

Diversos estudios demuestran la relación entre el acceso a la vivienda y la superación de la pobreza. Sin embargo, existe un rezago en el acceso a la vivienda digna en México y la falta de historial crediticio es una limitante para el acceso a créditos bancarios. El objetivo de la presente investigación es analizar los criterios de selección de crédito hipotecario y proponer un modelo de gestión de riesgos que permita a la banca financiar a un mayor número de personas en la adquisición o mejora de su vivienda. La estrategia metodológica se basa en técnicas de aprendizaje automático apoyadas en la ciencia de datos para crear un modelo predictivo del cumplimiento del crédito basado en características individuales. Los resultados muestran un modelo predictivo de discriminación de riesgo con una confiabilidad del 85% para créditos a la vivienda, lo cual permite ampliar la base potencial de personas susceptibles de acceder a financiamiento hipotecario. El derecho a una vivienda digna presenta un rezago importante en el país y hasta ahora los bancos al proponer un modelo predictivo de selección de riesgo hipotecario se da respuesta a la pregunta de investigación que refiere a las acciones que puede ejecutar la banca para resolver el problema de falta de acceso a vivienda digna. Los bancos pueden establecer sus criterios de selección de riesgo apoyados en la ciencia y analítica de datos y la aplicación de modelos predictivos de aprendizaje automático utilizando su amplia base de datos histórica.

Palabras clave: aprendizaje automático; criterios de selección de riesgo; movilidad social; vivienda digna.

Recibido: 07.02.23

Aceptado: 08.05.23

* Maestro. Autor corresponsal e investigador en la en la Universidad Panamericana. Facultad de Ciencias Económicas y Empresariales, Augusto Rodin 498, Ciudad de México,03920. México Email: 0230199@up.edu.mx, ORCID: <https://orcid.org/0000-0003-2558-157X>.

** Maestra. Autor corresponsal, Profesora e Investigadora en la Universidad Panamericana. Facultad de Ciencias Económicas y Empresariales, Augusto Rodin 498, Ciudad de México,03920. México. Email: lotorre@up.edu.mx, ORCID: <https://orcid.org/0000-0001-9297-4758>.

*** Doctora. Autor corresponsal, Profesora e Investigadora en la Universidad Panamericana. Facultad de Ciencias Económicas y Empresariales, Augusto Rodin 498, Ciudad de México,03920. México. Email: ateran@up.edu.mx, ORCID: <https://orcid.org/0000-0002-0240-5234>.

Mortgage credit: a predictive model of risk discrimination

Abstract

Several studies demonstrate the relationship between access to housing and overcoming poverty. However, there is a lag in access to decent housing in Mexico and the lack of credit history is a limiting factor for access to bank loans. The objective of this research is to analyze mortgage credit selection criteria and propose a risk management model that allows banks to finance a greater number of people in the acquisition or improvement of their homes. The methodological strategy is based on machine learning techniques supported by data science to create a predictive model of credit fulfillment based on individual characteristics. The results show a predictive model of risk discrimination with a reliability of 85% for housing loans, which allows expanding the potential base of people susceptible to access mortgage financing. The right to decent housing presents a significant backlog in the country and so far the banks, by proposing a predictive model of mortgage risk selection, provide an answer to the research question that refers to the actions that can be taken by the banks to solve the problem of lack of access to decent housing. Banks can establish their risk selection criteria supported by data science and analytics and the application of predictive machine learning models using their extensive database.

Keywords: decent housing; machine learning; risk selection criteria; social mobility.

1. Introducción

El acceso a la vivienda es un derecho de los individuos establecidos por los países miembros de la Organización de las Naciones Unidas (ONU), en el artículo 25 de la Declaración Universal de los Derechos Humanos (ONU, 1948), al establecer que “Toda persona tiene derecho a un nivel de vida adecuado que le asegure, así como a su familia, la salud y el bienestar, y en especial la alimentación, el vestido, la vivienda, la asistencia médica y los servicios sociales necesarios”.

Por lo tanto, los gobiernos y los participantes del mercado deben

procurar asegurar el ejercicio de ese derecho a la población. Los primeros pueden contribuir incrementando sus programas de garantía para personas de bajos ingresos, los desarrolladores y constructores ejecutando proyectos en que se habiliten viviendas en las zonas cercanas a servicios generales en lugar de la periferia, y los bancos e intermediarios financieros innovando en sus modelos de gestión de riesgo y con productos apropiados.

En México, este derecho no se ha cubierto, por lo que resulta de interés entender las posibles causas de la brecha negativa entre demanda y oferta

para proponer alguna alternativa de solución.

Además, en el país más de 3,250 mil viviendas tienen una carencia por materiales o servicios y existe un déficit de unidades de más de 14 millones. De 2013 al 2021, los bancos han contribuido solo con el 20% del total de 5.5 millones de acciones emprendidas en conjunto con organismos públicos y de fomento a la vivienda de interés social con valor inferior a \$ 860 mil pesos y su oferta de crédito se concentra en la modalidad de adquisición.

Lo anterior impide garantizar el derecho a una vivienda digna y el papel de las instituciones financieras es relevante para incidir favorablemente en el ejercicio del derecho al acceso a vivienda entre la población. Un cambio en los criterios actualmente empleados para la asignación de créditos puede ser una vía, pues, actualmente, uno de los principales obstáculos que los solicitantes de crédito hipotecario enfrentan en México es el cumplimiento de criterios tradicionalmente empleados por los bancos en la evaluación del riesgo, particularmente para quienes no han tenido créditos anteriores y carecen de historial crediticio.

La evaluación de riesgos es un proceso crucial para los bancos, pues, requieren tomar decisiones informadas sobre la concesión de préstamos y dicho proceso puede ser costoso y requerir mucho tiempo.

En ese contexto, la pregunta que guía la presente investigación es: ¿Cómo pueden los bancos favorecer el acceso a la vivienda digna para la población a través de un modelo de gestión de discriminación de riesgo? La cual permitirá el análisis de los criterios de selección de crédito hipotecario.

Para atender a esta interrogante, se propone un modelo de gestión de

riesgo de crédito hipotecario basado en algoritmos de aprendizaje automático, que emplean variables relacionadas con las características del individuo, pero que no están ligadas con comportamientos en créditos anteriores, y se presenta como una solución prometedora para facilitar el acceso a la vivienda.

Li y Zhao (2020) encontraron que el uso de aprendizaje automático puede mejorar la precisión de la evaluación de riesgos en un 14% en comparación con métodos tradicionales.

Este trabajo se integra por seis apartados. En el primero se aborda la introducción y se formula la pregunta de investigación; en el segundo, se revisa el marco teórico y la evolución de los criterios de discriminación de riesgo hipotecario. Posteriormente, en la tercera sección se aborda la situación actual de la vivienda y su financiamiento en México y la estrategia metodológica se establece en la cuarta sección. El modelo predictivo y sus resultados se muestran en el quinto apartado y por último en el sexto, se citan las conclusiones.

2. Evolución de los criterios la selección de riesgo de crédito hipotecario

Los criterios de selección de riesgo hipotecario que utilizan los bancos y otras instituciones financieras se basan en lo que denominan las cinco C's del crédito como lineamiento general para tomar decisiones de financiamiento y, a partir de ellas, emplean modelos paramétricos propios para seleccionar el riesgo que están dispuestos asumir. Las cinco C's del crédito consideran el carácter del solicitante, su capacidad de pago, el capital del que dispone, el colateral o garantía y las condiciones económicas

prevalcientes en su entorno (Treece & Tarver, 2021).

Para evaluar el carácter del solicitante, las instituciones de crédito utilizan el comportamiento histórico de crédito, su relación con otros acreedores, su historia laboral, referencias personales, credenciales y reputación (Peprah y Oten, 2017). En México, las agencias de información crediticia, como Buró de Crédito y Círculo de Crédito, son importantes proveedores de datos relacionados con el comportamiento de pagos del solicitante (Centro de Estudios Espinosa Yglesias, 2017).

La capacidad de pago se determina como proporción del ingreso disponible después de cumplir con otras obligaciones de pago (Baiden, n.d). La estabilidad del ingreso y su magnitud son factores críticos; típicamente el monto del pago correspondiente al crédito solicitado no debe ser superior al 30% del efectivo disponible para deuda. Las instituciones mexicanas suelen requerir comprobantes oficiales de nómina, honorarios por servicios prestados o estados de cuentas históricos de cuentas de cheques en que se verifican los flujos de entrada y salida de efectivo (Centro de Estudios Espinosa Yglesias, 2017).

El capital disponible representa el compromiso del solicitante de aportar recursos propios y en los créditos hipotecarios corresponde al nivel de enganche aportado por el demandante del crédito (Peprah y Oten, 2017). Los bancos, por lo general, solicitan al menos un compromiso del 10% de la adquisición o inversión; el colateral es el valor de una garantía tangible o real (en el hipotecario es el inmueble) que reduce la exposición del banco y servirá como fuente secundaria de pago en caso de falta de pago. Además de considerar la salud financiera del solicitante, las

instituciones de crédito toman en cuenta otras condiciones financieras como el estado general de la economía, las tasas de interés y el destino del crédito (Baiden, 2011).

Una vez considerados los factores anteriores, los bancos utilizan modelos paramétricos para obtener una resolución de aceptación o rechazo. Aunque los modelos en México no son públicos, se considera que los factores influyentes en el resultado, de acuerdo con el puntaje de crédito FICO utilizando en Estados Unidos son: historial crediticio (35%), monto de deuda contratada (30%), antigüedad de la historia crediticia (15%), diversidad de fuentes de financiamiento (10%) y número de solicitudes de crédito abiertas en fecha reciente (10%) (Treece & Tarver, 2021). En resumen, el criterio de selección de riesgo es la capacidad de pago demostrable, un historial de crédito sano, la capacidad de aportar recursos propios o capital y la existencia de una fuente secundaria de pago de fácil realización (Baiden, 2011).

Los avances tecnológicos derivados del acceso a internet, sistemas computacionales y telefonía celular han aumentado exponencialmente la cantidad de información que las empresas generan y tienen a su disposición cada día. El análisis del riesgo crediticio es un ejemplo de la modelación computacional que ha renacido con los avances en los algoritmos de aprendizaje automatizado, gracias a la gran cantidad de datos y a la complejidad de las posibles relaciones que se pueden establecer entre las características de los consumidores y sus transacciones (Khandani et al., 2010).

El concepto de aprendizaje automático (o machine learning) en el ámbito de la literatura de ciencias computacionales se refiere a un

conjunto de algoritmos que buscan encontrar patrones contenidos dentro de enormes cantidades de bases de datos (Khandani et al, 2010). Su desarrollo ha ido de la mano del uso de Big Data, o datos masivos, caracterizados por tres aspectos: volumen (en la cantidad de bytes), velocidad (de creación, de utilización, en tiempo real) y variedad (diferentes tipos de datos) (Guedea-Noriega y García-Sánchez, 2019).

Cortez y Rocha (2010) muestran cómo el empleo de aprendizaje automático puede mejorar la precisión en la evaluación de riesgos de crédito en Portugal; en tanto, García-Solarte et al, (2017) al hacer una revisión exhaustiva de los estudios que utilizan aprendizaje automático para evaluar riesgos de crédito discuten las fortalezas y debilidades de las diferentes técnicas y su aplicabilidad en diferentes contextos. En Chipre, Papadopoulus et al, (2015) desarrollaron modelos predictivos de riesgo crediticio empleando técnicas de minería de datos y los compararon con los modelos tradicionales de evaluación de riesgos, obteniendo resultados satisfactorios.

La discusión respecto a cómo el aprendizaje automático puede mejorar la gestión de riesgos de crédito en el sector financiero y sus implicaciones para la toma de decisiones es abordada por Esperón-Moreno y García-Sánchez (2019).

La literatura sobre el análisis del riesgo crediticio muestra evidencias de los beneficios del uso de algoritmos de clasificación que hacen uso de grandes cantidades de información, adicional a la que los puntajes de crédito tradicionales emplean. Los avances tecnológicos hacen posible el uso de diversos tipos de datos que se obtienen cada día en grandes cantidades, con el objetivo de

identificar a los clientes morosos y no morosos.

La combinación de avances tecnológicos con modernas técnicas de análisis de información usando minería de datos permite una toma de decisiones más eficiente, lo que a su vez se puede traducir en una mayor rapidez para otorgar los créditos, con un menor costo y a una fracción mayor de la población, favoreciendo la inclusión financiera (Bazarbash, 2019).

De acuerdo con la información reportada por los bancos a la Comisión Nacional Bancaria y de Valores (CNBV), existen distintas modalidades de financiamiento al sector inmobiliario de acuerdo con el destino del crédito como son: adquisición de terreno para vivienda, adquisición de terreno y construcción, adquisición de vivienda nueva o usada, crédito para liquidez, mejoras, ampliaciones y/o remodelaciones, pago de pasivos hipotecarios (crédito sustitución) y crédito para ex empleados de instituciones financieras (CNBV, 2020).

3. Situación actual de la vivienda y su financiamiento en México

La vivienda es un factor clave para el ser humano. No es solo porque satisface las necesidades básicas de protección, seguridad, salud y bienestar de las personas, sino porque también responde a las necesidades de identidad, pertenencia, privacidad y armonía. Además, la vivienda tiene un impacto indirecto y directo en la satisfacción de otras necesidades fundamentales como la salud, la educación, el trabajo y la convivencia. Esto apoya la formación de bienes familiares, reduce las

desigualdades y permite mejorar la calidad de vida de la población.

Así lo reconoce la ONU dentro del Pacto Internacional de Derechos Económicos, Sociales y Culturales (ONU, 1966). Aún más, la Agenda 2030 para el Desarrollo Sostenible propone, en su objetivo 11.1 “De aquí a 2030, asegurar el acceso de todas las personas a viviendas y servicios básicos adecuados, seguros y asequibles y mejorar los barrios marginales” (ONU, 2015).

De esta forma, el acceso a la vivienda constituye un derecho humano reconocido a nivel mundial. En México, el derecho a la vivienda está contemplado como uno de los objetivos prioritarios del Programa Nacional de Vivienda 2021-2024, según se publicó en el Diario Oficial de la Federación (DOF, 4 de junio 2021):

La vivienda se considera como un instrumento para afrontar retos tales como la pobreza, la desigualdad e incluso el cambio climático. Sin embargo, ONU-Habitat (2018), estima que al menos 38.4% de la población mexicana habita una vivienda no adecuada (ONU-Habitat, 2018):

El fundamento normativo en materia de vivienda en México se ubica en la Constitución Política de los Estados Unidos Mexicanos, que establece en su artículo 4º. el derecho de todas las familias a disfrutar de una vivienda digna y decorosa. De ahí se desprende la Ley de Vivienda, donde la vivienda se establece como un área prioritaria para el desarrollo nacional, y garantiza que el Estado organizará e impulsará las actividades necesarias para cumplirlo (DOF, 4 de junio 2021).

El Programa Nacional de Vivienda 2019-2024 es el primero en incorporar los siete elementos de la vivienda adecuada

definidos por ONU-Habitat, buscando que todos los actores impulsen estos criterios en sus planes y programas. Toma como fundamento tres principios rectores del Plan Nacional de Desarrollo (DOF, 4 de junio 2021):

I. No dejar a nadie atrás, no dejar a nadie fuera, al colocar a las personas con mayor vulnerabilidad social en el centro de la estrategia

II. Honradez y honestidad, al proponerse incrementar la transparencia y reducir la corrupción en todos los procesos del sistema de vivienda

III. El mercado no sustituye al Estado, al promover la participación del sector privado y social bajo un marco institucional de certeza y claridad, que lo convierta en aliado corresponsable en el desarrollo de vivienda adecuada en el país.

Así, para lograr el objetivo de derecho a la vivienda, las estrategias están enfocadas hacia disminuir el número de viviendas en condiciones de rezago, tanto en las zonas urbanas como en las rurales (DOF, 4 de junio 2021).

Este objetivo de lograr el acceso a una vivienda digna para mejorar las condiciones de vida, ha sido motivo de políticas públicas en México y Latinoamérica. Ha transitado de un modelo de arriendo (1950) y autoconstrucción (1960), programas de vivienda centralmente planeados y apoyados por el Estado- mediante fuentes públicas y privadas-, y localización geográfica en áreas cercanas a los servicios, fuentes de empleo e infraestructura (hasta 1980); hacia uno en donde se han expandido los centros urbanos hacia la periferia y en el que existe una mayor participación de las instituciones financieras en el crédito a los individuos con ingresos medios y bajos, mediante subsidios y

garantías otorgados por organismos oficiales de apoyo a la vivienda (Imilan y Beswick, 2016).

Por otra parte, el sector vivienda tiene una estrecha vinculación con la economía. Por ejemplo, en 2015, concentró aproximadamente el 6% del producto interno bruto (PIB); en particular, la edificación de vivienda aportó alrededor del 66% del PIB de la construcción e incidió en 78 clases de actividad económica (Rodríguez y Morales, 2021).

Además, el Consejo Nacional de Población estimó que la demanda de nuevos hogares crecerá a un ritmo de 581 mil cada año entre 2010 y 2030 (Consejo Nacional de Vivienda [CONAVI], 2016), y existen casi 10 millones de hogares con ingresos inferiores a \$ 12,500 pesos mensuales con un rezago habitacional por materiales deteriorados (piso de tierra, techos o paredes con material de desecho) o precariedad de espacios (promedio de más de 2.5 residentes por cuarto o ausencia de excusado) (CONAVI, 2021).

Según datos del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL, 2018), en México, existe un rezago de más de 14 millones de viviendas, lo que representa el 44.7% de viviendas. Los estados que presentan mayor nivel de rezago están ubicados en el sursureste: Chiapas, Oaxaca, Guerrero, Veracruz y Campeche, y en estos estados se ubica la población indígena que reporta mayor índice de rezago.

En ese contexto, existe una segmentación por tipo de vivienda en México que se realiza de acuerdo con el valor del inmueble expresado en veces, unidades de medida y actualización (UMA) que es una referencia económica en pesos publicada por el Instituto Nacional de Estadística y Geografía (INEGI) para determinar la cuantía del pago de obligaciones y supuestos previstos en la ley (Olivares-Aguayo et al, 2021). Los tipos de vivienda y su valor en UMA se detallan en la Tabla 1.

Tabla 1
Tipo de vivienda en UMA's

Tipo	Múltiplo UMA	Valor económico.
Económica	Hasta 118	Hasta \$ 317,254
Popular	De 118 a 200	De \$ 317,255 a \$ 487,950
Tradicional	De 200 a 350	De \$ 487,951 a \$ 853,910
<i>Interés Social = Económica + Popular + Tradicional</i>	<i>Hasta 350</i>	<i>Hasta \$ 853,910</i>
Media	De 350 a 750	De \$ 853,910 a \$ 1,829,809
Residencial	De 750 a 1,500	De \$ 1,829,810 a \$ 3,659,617
Residencial Plus	Mayor a 1,500	Mayor a \$ 3,659,917

Fuente: Elaboración propia con base en Olivares-Aguayo et al, (2021).

El valor de la UMA en 2021 es de 89.62 pesos diarios y el múltiplo aplicable se calcula con la UMA mensual equivalente a \$ 2,439.74 (Olivares-Aguayo et al, 2021).

De 2013 a agosto 2021 las estrategias empleadas por el sector bancario han logrado aportar únicamente el 20% de las acciones (financiamiento, subsidios o garantías) para vivienda de

interés social de un total de 5,565,580 acciones (CONAVI, 2016).

De acuerdo con Rodríguez y Morales (2021), en las ciudades con más de 100,000 habitantes, las viviendas compradas en 2016 representaron 52.3% del total, la mayoría estuvieron vinculadas con acceso al crédito y de éstas, el 71% se originaron con alguna institución financiera incluyendo los programas con participación de Instituto del Fondo Nacional de Vivienda para los Trabajadores (INFONAVIT) y el Fondo de la Vivienda del Instituto de Seguridad y Servicios Sociales del Estado (FOVISSSTE).

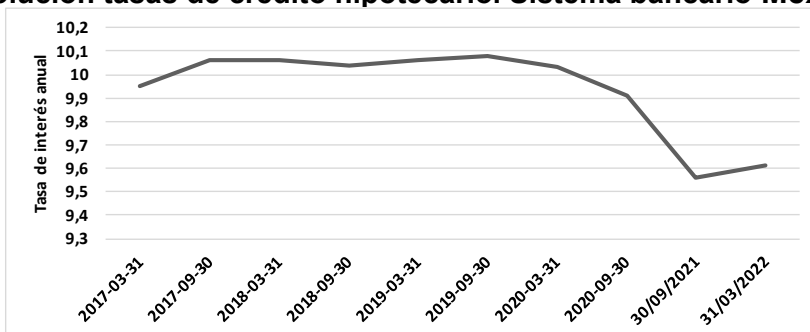
Entre 2010 y 2015 la banca comercial, el INFONAVIT y el FOVISSSTE, otorgaron en promedio 381,000 créditos cada año para adquisición de vivienda nueva respecto de 388,000 unidades promedio de vivienda nueva inscritas en el registro único de vivienda (RUV) (Sociedad Hipotecaria Federal [SHF], 2016). A pesar de ello, los avances de los sistemas de financiamiento para la vivienda medidos por su participación respecto al producto nacional bruto

siguen siendo relativamente pequeños; mientras en México se ubican en 10%, en economías más pequeñas como Bolivia o Costa Rica reportan 11% y 16% de participación respectivamente y en Chile se ubican con el 22% (Eloy, 2017).

Olivares-Aguayo et al, (2021) mencionan que entre las restricciones que enfrenta una familia para acceder a un crédito hipotecario se encuentran: la volatilidad de las tasas de interés, la insuficiencia de recursos para cubrir las necesidades básicas y la ausencia de salud financiera al no contar con disciplina de ahorro ni saber determinar su capacidad de pago; así, concluyen que un individuo con un ingreso mensual entre \$ 2,103 pesos y \$ 6,309 pesos no tiene la capacidad de pago necesaria para hacer frente a una hipoteca.

En contraste, la tasa bancaria fija actual para hipotecas a plazos de 15 y 20 años se encuentra en un nivel inferior al 9.70% anual (Gráfico 1) y la incursión de otros intermediarios como las sociedades financieras de objeto múltiple han intensificado la competencia por el mercado hipotecario.

Gráfico 1
Evolución tasas de crédito hipotecario. Sistema bancario México



Fuente: Elaboración propia con base en Banco de México (<https://www.banxico.org.mx/PortalTranspCompSistFin/>).

Por otra parte, la industria inmobiliaria se comporta de manera cíclica a contra flujo del comportamiento de los mercados financieros. Cuando las tasas de interés bajan, las ventas suben, la construcción se incrementa y los precios de los bienes inmuebles se elevan (Imilan y Beswick, 2016). Esto origina que los precios de los activos financieros bajen, se presenta una sobre oferta de bienes inmuebles, las tasas de interés suben, bajan las ventas, la construcción cae, los precios de los bienes inmuebles se contraen, suben los rendimientos sobre los activos financieros, se contrae la oferta de bienes, bajan las tasas de interés (Imilan y Beswick, 2016); y así, se repite el ciclo.

La evolución reciente indica una menor participación de créditos para vivienda de interés social respecto del total y, por lo tanto, representa una oportunidad para aquellos bancos que utilicen información no estructurada en sus modelos de evaluación de riesgo que les permita identificar cierto perfil y/o patrones de consumo de los solicitantes,

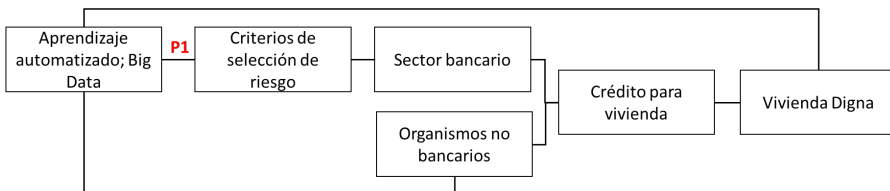
para calcular la capacidad de pago de los individuos y familias que operan en el sector informal, que no tienen acceso a crédito bancario, y que son demandantes de vivienda social.

En resumen, en México existe un déficit habitacional del orden de 9.8 millones de viviendas en 2020, cercano al 22.5% del total de hogares del país (INEGI, 2020). El precio promedio de una vivienda es alto; al primer trimestre 2021 se ubicó en \$ 2.6 millones (CONAVI, 2021) y el 41% presentan al menos una deficiencia en su estructura, servicios básicos o equipamiento (Secretaría de Desarrollo Agrario, Territorial y Urbano [SEDATU], 2018).

4. Estrategia metodológica

La presente investigación es de tipo cuantitativa y se utiliza un modelo predictivo a partir de un análisis de los datos y las relaciones entre las variables (Creswell, 2014) la cual que busca mostrar la factibilidad de la proposición establecida en el Diagrama 1.

Diagrama 1
Modelo predictivo



Fuente: elaboración propia

La disponibilidad de créditos para vivienda en México proviene principalmente de dos fuentes: el sector bancario y los organismos no

bancarios. Para la población que recibe el crédito, tener acceso a esos recursos le permite ya sea hacerse de un bien inmueble o bien mejorar las condiciones

de su vivienda actual, contribuyendo al derecho humano del acceso a vivienda digna.

Para que el sector bancario pueda aumentar su participación en este mercado, puede hacer cambios en sus criterios de selección de riesgo, integrando información procesada a través de técnicas de aprendizaje automatizado y minería de datos, que representan un cambio innovador a los criterios tradicionalmente utilizados. Los resultados de estos cambios en criterios se pueden traducir en un mayor número de clientes recibiendo créditos y una mayor variedad de créditos, no solo de adquisición sino también para mejoras de vivienda.

Los modelos de aprendizaje automatizados requieren grandes cantidades de datos, por lo que toda la información generada por los créditos ya otorgados se emplea para retroalimentar el modelo y mejorar su capacidad de predicción, por lo tanto, se realizó una

búsqueda de información pública de solicitantes de crédito y su posterior desempeño de pago en México y al no encontrar datos abiertos se utilizaron dos bases de públicas que corresponden a un banco comercial chino. La primera, con la información personal no sensible de quienes solicitaron una tarjeta de crédito y la segunda, con el comportamiento histórico de diversos deudores de tarjeta de crédito del mismo banco en un mes determinado. Ambas bases de datos se relacionan mediante el campo de número cliente (se encuentran disponibles en www.kagle.com/code/rikdifos/ Credit-card-approval-prediction-using-ml).

Utilizando el código en lenguaje Python desarrollado por Song (2019) sobre la base de datos histórica que registra el estatus de pago de los deudores y que incluye, entre otros datos, el periodo en que ha estado vigente la tarjeta en un mes determinado en la Tabla 2, con 1,048,575 renglones.

Tabla 2
Comportamiento histórico de tarjeta de crédito

Nombre del campo	Contenido	Observaciones
ID	Identificador de número de cliente	1,048,575 registros. Cada registro identifica el comportamiento de un cliente en un mes determinado.
Meses de balance	Número de meses transcurridos desde que se otorgó el crédito. Cuenta hacia atrás a partir de la fecha en que se extrajeron los datos.	
Status	Al corriente, con retraso de 1-29 días; de 30 a 59; de 60 a 89; de 90 a 119; de 120 a 149; de 150 en adelante; sin crédito.	

Fuente: Elaboración propia con base en Song, X. (2019).

Se define como variable respuesta “crédito malo” aquel cuyo número de pagos registrados con retraso es mayor al número de pagos que el cliente debió realizar desde que recibió y utilizó la tarjeta de crédito hasta la fecha en que se

extrae la información. Posteriormente, se relaciona este conjunto de datos con la de los aplicantes de crédito correspondientes a 90,085 personas (cuadro 2) y se obtienen así 9,709 registros que se interceptan por número de cliente en ambas bases.

Cuadro 2 Solicitud de aplicantes de crédito

Nombre del campo	Contenido	Observaciones
ID	Identificador de número de cliente	439 mil registros
Género	Masculino o Femenino	
Indicador de auto propio	Verdadero o Falso	
Indicador de casa propia	Verdadero o Falso	
Número de hijos sin importar si viven con sus padres o no.	De 0 a 19	
Ingreso anual del aplicante	En unidades monetarias	
Origen del ingreso	Trabajo, comercio, trabajador del estado, pensionado, estudiante, otro.	
Tipo de educación recibida	Primaria, Secundaria/especial, educación media superior incompleta, educación media superior completa, licenciatura , otra.	
Forma de vida	Casa, departamento, casa o departamento con padres, departamento municipal, departamento rentado, oficina departamento, otra.	
Estado civil	Casado, soltero, viudo, separado, otro	
Fecha de nacimiento.	Edad en días contados a partir de la fecha actual hacia atrás.	
# de días empleado o autoempleado.	Antigüedad laboral en días. Cuenta hacia atrás a partir de la fecha actual. Si el número es positivo, el aplicante no tiene empleo.	
Indicador de celular	Si / No	
Indicador de teléfono de trabajo	Si / No	
Indicador de teléfono fijo en el lugar que habita.	Si / No	
Indicador de email	Si/ No	
Tipo de ocupación	Trabajador, otro.	
Cantidad de miembros de la familia que viven juntos.	Número de 1 en adelante.	

Fuente: Elaboración propia con base en Song (2019).

A continuación, mediante el software de aprendizaje automático denominado Orange se normalizan los datos a $\mu = 0$ y $\sigma^2 = 1$ y se prueban distintos algoritmos de clasificación: K-Vecinos cercanos (KNN), árboles de decisión (*Decision Tree*), bosque aleatorio (*Random Forest*) y Naive Bayes; y se obtiene un modelo predictivo. La métrica ocupada para evaluar la capacidad predictiva de cada

algoritmo de clasificación denominada “Accuracy”, denota con un nivel de 98.9% que *Random Forest* es el mejor modelo (Tabla 3) y utiliza las siguientes siete variables como las más relevantes para la predicción: meses de balance, status, número de días con empleo o autoempleo, tipo de educación recibida, estado civil, indicador de casa propia e indicador positivo de auto propio.

Tabla 3
Comparativa bajo criterio CA (Accuracy) de modelos predictivos -datos completos-

Model	AUC	CA	F1	PRECISION	RECALL
kNN	0.999	0.984	0.984	0.984	0.984
Tree	0.988	0.979	0.979	0.979	0.979
Random Forest	1.000	0.989	0.989	0.989	0.989
Naive Bayes	0.959	0.861	0.859	0.869	0.861

Fuente: Elaboración propia utilizando software Orange con base en conjunto de datos original.

Para evaluar el modelo predictivo de evaluación de solicitudes de crédito para vivienda, se elimina de la base de datos de los solicitantes de crédito el campo relativo el indicador de casa propia, con el fin de asegurar que no la posee, y se excluyen de la base del comportamiento de crédito de los deudores de tarjeta de crédito, las columnas que muestran los meses de balance y el estado relacionado con sus antecedentes de cumplimiento de obligaciones de crédito.

De esta forma, esas variables no intervienen para la construcción de una base de datos ajustada y un modelo

predictivo ajustado. Es oportuno resaltar que típicamente la cartera morosa en tarjeta de crédito es superior a la cartera hipotecaria, por lo que utilizar la base de datos de solicitantes de tarjeta de crédito se considera pertinente.

Con la base de datos ajustada se ejecutan los algoritmos de clasificación previamente citados más Adaboost, que es un algoritmo utilizado para mejorar la precisión de los modelos predictivos mediante la combinación de múltiples aprendices débiles (se realizaron 200 estimaciones). Al comparar los modelos por su "Accuracy", Adaboost se ubica como el mejor con un nivel 85 % (Tabla 4).

Tabla 4
Comparativa bajo criterio CA (Accuracy) de modelos predictivos datos ajustada

Model	AUC	CA	F1	PRECISION	RECALL
kNN	0.710	0.661	0.661	0.662	0.661
Tree	0.871	0.763	0.756	0.776	0.763
Random Forest	0.897	0.794	0.791	0.799	0.794
Naive Bayes	0.519	0.542	0.437	0.516	0.542
AdaBoost	0.951	0.850	0.849	0.852	0.850

Fuente: Elaboración propia utilizando software Orange con base en conjunto de datos ajustada

La información más relevante que ocupa Adaboost para la construcción del modelo predictivo usando un criterio de chi cuadrada es la contenida en los siguientes cinco campos: número de días con empleo o autoempleo, tipo de educación recibida, estado civil, indicador de auto propio y origen del ingreso.

5. Modelo predictivo de gestión de riesgo hipotecario. Resultados

A través de la técnica Adaboost de aprendizaje automático supervisado basada en árboles de decisión y la combinación de diversos aprendices débiles, se encontró un modelo de predicción con una precisión del 85% para determinar préstamos hipotecarios, buenos y/o malos.

Dado que el comportamiento de los préstamos de tarjetas de crédito con retrasos suele ser peor que el de la cartera hipotecaria vencida, el modelo de predicción muestra que los criterios

de selección de riesgos basados en un volumen masivo de información y análisis de datos son viables para evaluar las solicitudes de crédito de la población aún y cuando en el conjunto de datos no se cuente con información histórica de créditos obtenidos ni con un indicador de posesión de casa propia. Este análisis es relevante porque el acceso a una vivienda digna mediante crédito hipotecario es posible para los aplicantes que lo soliciten, a los bancos que estén dispuestos a adoptar el modelo predictivo de evaluación de riesgo propuesto y los bancos que lo adopten tendrán incentivos para hacerlo, pues, pueden ganar cuota de mercado al proporcionar crédito a personas que no tienen acceso a financiamiento hoy en día.

De 2013 a agosto 2021, las estrategias empleadas por el sector bancario han logrado aportar únicamente el 20% de las acciones (financiamiento, subsidios o garantías) para vivienda de interés social (Tabla 5).

Tabla 5
Número de acciones para vivienda de interés social por organismo¹

ORGANISMO	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total general
BANCA (CNBV)	46,309	50,147	34,031	123,518	176,067	196,890	196,830	57,150	126,847	1,107,789
BANJERCITO	3,091	1,667	1,379	1,048	737	330	186	224	161	8,823
CFE	463	351	339	166	465	1,516	2,296	215		5,811
CONAVI	162,098	247,544	197,708	140,315	97,699	104,413	33,592	10,147	1,101	994,617
FONHAPO	33,851	78,995	102,056							214,902
FOVISSSTE		33,476	37,030	34,612	26,677	26,053	26,685	25,520	21,210	231,263
HABITAT MEXICO	1,801	1,328								3,129

Cont... tabla 5

ORGANISMO	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total general
IMEVIS	1,660	2,845	3,302							7,807
INFONAVIT	337,344	336,707	343,229	318,824	335,479	317,556	291,583	283,141	189,356	2,753,219
INVI	8,547	13,658	14,472	16,847	10,682	13,955	10,784	12,103	12,247	113,295
ISSFAM	2,958	2,675	3,010	2,822	2,468	2,325	1,245			17,503
ISSSTELEON	634									634
IVEM	69	74	71							214
PDZP SEDESOL	80,273	2,774								83,047
SHF (FONDEO)	5,533	7,544	241	77	53	1,878	5,200	3,001		23,527
Total general	684,631	779,785	736,868	638,229	650,327	664,916	568,401	491,501	350,922	5,565,580

¹ Incluye financiamiento para adquisición de terreno, de vivienda nueva o usada, para la construcción de vivienda propia, crédito para liquidez, crédito para mejoras, ampliaciones y remodelaciones. Fuente: Elaboración propia con base en CONAVI (2016).

Los organismos no bancarios continúan siendo los principales proveedores de financiamiento para el sector de la población que busca este tipo de vivienda.

De acuerdo con datos del censo de población y vivienda 2020 (INEGI, 2020), en México existen 35 millones 219 mil hogares de los cuales casi el 10% presentan una carencia por tener piso de tierra, no contar con drenaje o con sanitario (Consejo Nacional de Evaluación de la Política de Desarrollo Social [CONEVAL], 2020). Esto se evidencia al observar que las diferentes modalidades de financiamiento se concentran para la adquisición de vivienda nueva o usada y se originan pocas operaciones para financiar mejoras y adecuaciones a los hogares, adquisición de terrenos para vivienda, crédito a la construcción.

Complementar la oferta de financiamiento para las modalidades

citadas es una vía para impulsar el acceso a vivienda digna. Por tanto, un enfoque hacia modalidades diferentes de financiamiento requiere de criterios de selección de riesgo diferentes. De aquí se deriva la propuesta principal del presente documento: emplear modelos de discriminación de riesgo soportados en herramientas de aprendizaje automático para incrementar el impacto positivo que el sector bancario tiene sobre el acceso a vivienda digna.

Es a través de un cambio en los criterios tradicionales empleados para selección de riesgo que se puede atender un mayor número de individuos o familias que con los criterios actuales son rechazados cuando solicitan un crédito hipotecario o un préstamo para mejorar su vivienda. Se propone que, para evaluar la capacidad de pago de un solicitante, los bancos utilicen modelos de aprendizaje automatizado supervisado mediante el cual puedan

predecir su comportamiento de pago, aun sin contar con comprobantes de ingreso oficiales o estados de cuenta, y aun cuando el solicitante no tenga experiencia crediticia previa.

Para hacerlo, las instituciones pueden ocupar la amplia base de datos históricos de sus clientes actuales, quienes al solicitar un crédito hipotecario facilitan información como edad, género, actividad, domicilio, código postal, nivel de educación, enganche, valor estimado de la propiedad, declaran un cierto nivel de ingreso, estado civil, dependientes económicos, entre otros.

Los resultados obtenidos se alinean con los hallazgos de estudios existentes en la literatura que han utilizado algoritmos de aprendizaje automatizado en la determinación del nivel de riesgo crediticio de los clientes (Chen y Wang, 2019; Huang y Zhang, 2019; Li y Zhang, 2020; Miao y Huang, 2020). Mundra et al, (2019), por ejemplo, utilizaron una base de datos de 3,000 registros y 24 atributos a la que aplicaron cuatro diferentes metodologías de clasificación: Máquinas de Vectores de Soporte, K-Vecinos más cercanos, Regresión Logística y Redes Neuronales artificiales. Sus conclusiones indican que las Redes Neuronales son el método que genera mayor precisión (Accuracy), pero todos los métodos evaluados tuvieron niveles de precisión en la clasificación cercana o superiores a 80%.

Por su parte, Addo et al, (2018) emplearon red elástica (regresión logística con regularización), un bosque aleatorio, un modelado de aumento de gradiente y un enfoque de red neuronal con cuatro complejidades diferentes para determinar cuál de ellos ofrece una mejor clasificación de empresas solicitantes de crédito respecto a su solvencia. Como variables predictoras

emplearon información proveniente de los Estados Financieros de las empresas. Independientemente de cuál método resultó más adecuado, es relevante mencionar que entre sus conclusiones mencionan la importancia de contar con información de calidad y el uso de diferentes métodos con diferentes métricas de desempeño para elegir el más adecuado.

Un ejemplo más lo presentan Khandani et al, (2010), quienes combinaron información tradicionalmente empleada para el análisis de riesgo (tal como las razones de deuda a ingreso) con información de las transacciones bancarias de los consumidores, y lograron un incremento considerable en la capacidad de predicción de su modelo. Los considerables niveles de predicción logrados con los métodos anteriormente mencionados confirman la conveniencia de su uso, y la posibilidad de emplear diferentes atributos o características de los solicitantes, adicionales a las tradicionalmente empleadas.

Con el empleo de la minería de datos, modelos predictivos y aprendizaje automático, los bancos podrían otorgar financiamiento para nuevas hipotecas, pero principalmente para adecuaciones, terrenos, autoconstrucción, mejoras y renovaciones, con lo que se desarrollaría un subsegmento de mercado sin necesidad de contar con un colateral.

Las instituciones financieras en todo el mundo, y particularmente en México, tienen una ventana de oportunidad de hacer uso de los avances tecnológicos en el manejo de grandes volúmenes de datos para innovar técnicas de clasificación de riesgo, en beneficio tanto del crecimiento de su propio negocio como del mejoramiento de las condiciones de vida de la sociedad.

6. Conclusiones

El derecho a una vivienda digna presenta un rezago importante en el país y hasta ahora los bancos contribuyen poco a mejorarlo. Sus criterios de selección de riesgos representan una restricción para los individuos y familias que no tienen una actividad formal, historial crediticio o la documentación legal necesaria para constituir una hipoteca. Además, las modalidades de crédito a la vivienda se concentran en la adquisición de unidades nuevas o usadas, pero no fluye crédito para la remodelación, construcción o equipamiento.

Al proponer un modelo predictivo de selección de riesgo hipotecario en el que el rol de los bancos se potencie, se da respuesta a la pregunta de investigación que refiere a las acciones que puede ejecutar la banca para resolver el problema de falta de acceso a vivienda digna.

Los bancos pueden establecer sus criterios de selección de riesgo apoyados en la ciencia y analítica de datos y la aplicación de modelos predictivos de aprendizaje automático utilizando su amplia base de datos histórica. Con ello ampliarán la base de clientes potenciales y si además desarrollan modalidades de financiamiento complementarias a la adquisición, impactarán de manera favorable a las personas, al mercado, a los grupos de interés relacionados con el sector y ganarán en participación de mercado. Existe evidencia positiva en otras latitudes de la pertinencia de utilizar modelos de Big Data análisis para el otorgamiento de crédito.

Referencias bibliográficas

Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using

machine and deep learning models. *Risks*, 6(2), 38.

Baiden, J. (2011). *The 5c's of credit in the lending industry*. <http://ssrn.com/abstract=1872804> <https://doi.org/10.2139/ssrn.1872804>

Bazarbash, M. (2019). *Fintech in financial inclusion: machine learning applications in assessing credit risk*. International Monetary Fund. <https://doi.org/10.5089/9781498314428.001>

Blavy, R., & Souto, M. R. (2009). Estimating Default Frequencies and Macroeconomic Linkages in the Mexican Banking Sector, IMF Working Papers, (109). <https://doi.org/10.2139/ssrn.1413001>

Centro de Estudios Espinosa Yglesias. (2017). *El Sistema Financiero Mexicano . Diagnóstico y recomendaciones*. Ciudad de México.

Chen, Y., & Wang, T. (2019). Credit scoring based on artificial intelligence and deep learning. *Journal of Business Research*, 98, 365-380. <https://doi.org/10.1016/j.jbusres.2018.12.010>

Comisión Nacional Bancaria y de Valores. (10 de marzo de 2020). CNBV. <https://portafolioinfoctos.cnbv.gob.mx>

Conavi (2016). Sistema nacional de información e indicadores de vivienda. Conavi. <http://sniiv.conavi.gob.mx/>

Conavi (2021). Nota Metodológica. Cuantificación de la Población potencial y objetivo, Programa de Vivienda Social Clave presupuestaria S177

CONAVI. (2021). Precios de la vivienda en México. <https://www.gob.mx/conavi/articulos/precios-de-la-vivienda-en-mexico>

- Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) (2018). *Principales retos en el ejercicio del derecho a la vivienda digna y decorosa*. Ciudad de México: CONEVAL.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) (2020). *Índice de Rezago Social 2020*: CONEVAL
- Cortez, P., & Rocha, M. (2010). Credit scoring for loan granting: A case study. *Expert Systems with Applications*, 37(1), 144-154. <https://doi.org/10.1016/j.eswa.2009.03.009>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Díaz, S., González D., Borja, M.F., Puga, D. (2021). *Actualización del Rezago habitacional. Censo de Población y Vivienda 2020*. https://sniiv.conavi.gob.mx/doc/analisis/2021/Calculo_Rezago_Habitacional_2020.pdf
- Eloy, C. M. (2017). News from Latin America and the Caribbean; some facts and figures on the mortgage banking sector. *Housing Finance International*, 14–15.
- Esperón-Moreno, M. F., & García-Sánchez, T. (2019). The impact of machine learning on credit risk management. *European Journal of Management and Business Economics*, 28(3), 292-303. <https://doi.org/10.1108/EJMBE-03-2019-0037>
- García-Solarte, J. A., Muñoz, L. A., & Sánchez, G. M. (2017). Credit risk evaluation using machine learning methods: A review. *Journal of Business Research*, 77, 246-257. <https://doi.org/10.1016/j.jbusres.2017.03.005>
- Guedea-Noriega, H. H., & García-Sánchez, F. (2019). Semantic (Big) data analysis: An extensive literature review. *IEEE Latin America Transactions*, 17(05), 796-806. <https://doi.org/10.1109/TLA.2019.8891948>
- Huang, Y., Wu, H., Xu, G., & Zhang, J. (2019). Bank credit risk assessment based on convolutional neural networks. *Journal of Computational and Applied Mathematics*, 356, 227-235. [dhttps://doi.org/10.1016/j.cam.2018.11.008](https://doi.org/10.1016/j.cam.2018.11.008)
- Imilan, W., Olivera, P., & Beswick, J. (2016). Acceso a la vivienda en tiempos neoliberales: un análisis comparativo de los efectos e impactos de la neoliberalización en las ciudades de Santiago México y Londres. *Revista INVI*, 31(88), 163–195. <https://doi.org/10.4067/S0718-83582016000300006>
- Instituto Nacional de Estadística y Geografía- INEGI. (2021). Encuesta Nacional de Ingreso y Gasto de los Hogares. <https://www.inegi.org.mx/programas/enigh/nc/2020/>
- Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767-2787. <https://doi.org/10.1016/j.jbankfin.2010.06.001>
- Li, J., & Zhang, G. (2020). An empirical study on credit risk evaluation of SMEs based on machine learning algorithm. *International Journal of Machine Learning and Cybernetics*, 11(2), 317-326. <https://doi.org/10.1007/s13042-019-01011-8>
- Li, L., & Zhao, Z. (2020). Credit risk evaluation model based on machine learning. *Journal of Physics: Conference Series*, 1637(1), 012153. <https://doi.org/10.1088/1742-6596/1637/1/012153>

- Miao, R., & Huang, L. (2020). A deep learning approach to credit scoring with imbalanced data. *Journal of Business Research*, 114, 124-136. <https://doi.org/10.1016/j.jbusres.2020.04.016>
- Mundra, S., Mundra, A., Sharma, A., Mohini, A., & Yadav, A. (2019). Analyzing Credit Defaulter Behavior for Precise Credit Scoring. *Int. J. Adv. Sci. Technol*, 28(12), 247-255.
- Olivares-Aguayo, H. A., Méndez-Molina, M., & Madrigal-Castillo, E. (2021). Salud financiera en créditos hipotecarios mexicanos. *Revista CEA*, 7(13), 1–31. <https://doi.org/10.22430/24223182.1530>
- ONU-Habitat (2018). *Vivienda y ODS en México*. Programa de las Naciones Unidas para los Asentamientos Humanos.
- Organización de las Naciones Unidas (2015). *Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible*. https://www.un.org/ga/search/view_doc.asp?symbol=A/70/L.1&Lang=S
- Organización de las Naciones Unidas-ONU (1948). *Declaración Universal de Derechos Humanos*. <https://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=spn>.
- Organización de las Naciones Unidas-ONU (1966). *Pacto Internacional de Derechos Económicos, Sociales y Culturales*. <https://www.ohchr.org/SP/ProfessionalInterest/Pages/CESCR.aspx>
- Papadopoulos, T., Martakos, D., & Goudas, T. (2015). Credit scoring using data mining for the Cyprus banking sector. *Journal of Systems and Information Technology*, 17(4), 332-347. <https://doi.org/10.1108/JSIT-09-2014-0053>
- Peprah, W. K., Agyei, A., & Oteng, E. (2017). Ranking The 5C's of credit analysis: Evidence from Ghana banking industry. *International Journal of Innovative Research and Advanced Studies*, 9, 78-80. https://www.ijiras.com/2017/Vol_4-Issue_9/paper_17.pdf
- Programa Nacional de Vivienda 2021-2024 (2021) Diario Oficial de la Federación (DOF). http://www.dof.gob.mx/nota_detalle.php?codigo=5620388&fecha=04/06/2022
- Rodríguez Zamora, C., & Morales Ramírez, A. (2021). La duración en el mercado de vivienda nueva construida por desarrolladores en México. *Trimestre Económico*, 88(350), 551–582. <https://doi.org/10.20430/ete.v88i350.700>
- Secretaría de Desarrollo Agrario, Territorial y Urbano- SEDATU. (2018). Diagnóstico Nacional de Vivienda. Secretaría de Desarrollo Agrario, Territorial y Urbano. https://www.gob.mx/cms/uploads/attachment/file/373425/Diagnostico_Nacional_de_Vivienda.pdf
- Sociedad Hipotecaria Federal (2016). Dirección de Estudios Económicos de la Vivienda. Demanda de la vivienda 2016. <https://doc.shf.gob.mx/estadisticas/EstudiosVivienda/Documents/demanda%202016%20impresa.pdf>
- Song, X. (2019). Re: Credit Card Approval Prediction [Publicación en webpage de Kaggle, sitio de data analysis]. Kaggle. <https://www.kaggle.com/datasets/rikdifos/credit-card-approval-prediction>
- Treece, K., & Tarver J. (24 de mayo 2022). Understand The 5 Cs of Credit Before Applying for A Loan. *Forbes Advisory*. <https://www.forbes.com/advisor/credit-score/5-cs-of-credi>