

**PROPUESTAS PARA LA PREDICCIÓN DEL SOBREENDEUDAMIENTO EN  
HOGARES DE CHILE MEDIANTE EL USO DE UN MODELO HÍBRIDO QUE  
MEZCLA "ARTIFICIAL NEURO FUZZY INFERENCE SYSTEM" Y MODELO  
PROBIT**

Tesis de Grado presentado por

**Nicole Karina Astudillo Fingerhut**

como requisito parcial para optar al título de  
**Ingeniera Civil Industrial**

y al grado de

**Magíster en Ciencias de la Ingeniería Industrial**

Profesor Referente: Dr. Werner Kristjanpoller Rodríguez

Profesor Coreferente Interno: Dr. Javier Scavia

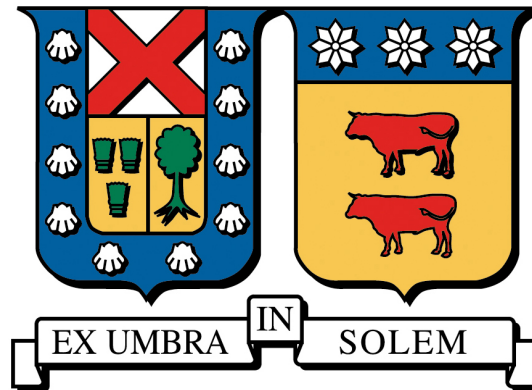
Profesor Coreferente Externo: Dr. Marcelo Villena

—

NOVIEMBRE 2018



UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA  
DEPARTAMENTO DE INDUSTRIAS



**PROPUESTAS PARA LA PREDICCIÓN DEL SOBREENDEUDAMIENTO EN  
HOGARES DE CHILE MEDIANTE EL USO DE UN MODELO HÍBRIDO QUE  
MEZCLA "ARTIFICIAL NEURO FUZZY INFERENCE SYSTEM" Y MODELO  
PROBIT**

Tesis de Grado presentado por

**Nicole Karina Astudillo Fingerhut**

como requisito parcial para optar al título de  
**Ingeniera Civil Industrial**

y al grado de

**Magíster en Ciencias de la Ingeniería Industrial**

Profesor Referente: Dr. Werner Kristjanpoller Rodríguez

Profesor Coreferente Interno: Dr. Javier Scavia

Profesor Coreferente Externo: Dr. Marcelo Villena

—

VALPARAÍSO, NOVIEMBRE 2018



TITULO DE LA TESIS:

**PROPUESTAS PARA LA PREDICCIÓN DEL SOBREENDEUDAMIENTO EN HOGARES DE CHILE MEDIANTE EL USO DE UN MODELO HÍBRIDO QUE MEZCLA "ARTIFICIAL NEURO FUZZY INFERENCE SYSTEM" Y MODELO PROBIT**

AUTOR:

**Nicole Karina Astudillo Fingerhut**

TRABAJO DE TESIS, presentado en cumplimiento parcial de los requisitos para el Grado de Magíster en Ciencias de la Ingeniería Industrial y de Ingeniera Civil Industrial de la Universidad Técnica Federico Santa María.

Dr. Werner Kristjanpoller Rodríguez .....

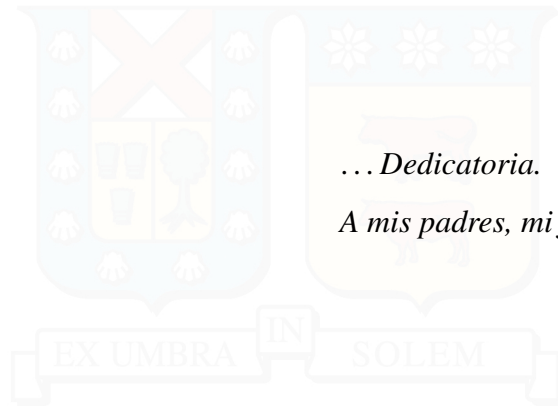
Dr. Javier Scavia .....

Dr. Marcelo Villena .....

---

VALPARAÍSO, Chile. NOVIEMBRE 2018





*...Dedicatoria.*

*A mis padres, mi familia y mi pareja...*



## AGRADECIMIENTOS

Gracias a mis padres, gracias Jorge y Karen por ser mis padres. Fueron, son y serán siempre mi apoyo y motivación para vencer las dificultades que trae la vida. Me hicieron fuerte y me entregaron la base para ser la mujer que soy hoy. Papá, gracias por darme todo lo que necesitaba para desarrollarme y avanzar, lograste que nunca me faltara algo y buscaste siempre tener todo lo que pedía más allá de las cosas materiales. Mamá gracias por haberte preocupado incondicionalmente siempre por mi educación y apoyarme cuando me sentía débil. Gracias a ambos por combinar en mí lo mejor de ustedes. Los amo infinitamente y siempre serán los mejores padres que pude haber tenido.

Gracias a mi abuela Cristina que estuvo presente todos estos años y mis abuelos ausentes que se mantuvieron presente en mi mente. Que alegría me da que puedas estar para compartir este gran momento conmigo, que tu salud y tu luz sigan presente por cuanto tiempo sea posible.

Gracias a mi hermana Natalie, espero algún día poder guiarte y apoyarte por esta entretenida, esforzada pero siempre gratificante vida universitaria. Aunque pareciera que no tenemos nada en común más que nuestros papas, espero que sepas que siempre estaré si me necesitas.

Gracias a la familia que la vida me entregó, este logro no es solo mío sino también de todos ustedes.

A la familia que yo elegí, a mi compañero de vida, gracias Miguel por tu infinito apoyo, tus palabras de aliento y tu eterna paciencia. Tu amor me hace bien. Me enseñaste que con amor se puede cambiar y que por amor la gente se apoya, se cae y se levanta junta. Fuiste la luz que iluminó el camino oscuro que enfrenté poco antes de este momento y demostraste ser mucho más que una buena pareja, demostrarte ser mi mitad perfecta. No tendré nunca palabras suficientes para explicarte lo mucho que me ayudaste a terminar este largo proceso que costó mucho, al cual le apostamos mucho pero que hoy nos entrega esta enorme felicidad. Comparte conmigo hoy y siempre la alegría y los frutos que nos traerá la vida.

A todos y todas esas personas que se cruzaron en mi vida durante estos años, con las

---

que pude compartir y aprender. A mis compañeros de carrera, tanto en el pregrado como magister. A esos profesores que fueron apareciendo y ayudando a construir y hacer realidad este sueño de ser Ingeniera.





---

## RESUMEN EJECUTIVO

Este trabajo propone un mejoramiento en la predicción del sobreendeudamiento de los hogares en Chile, utilizando para ello un innovador modelo híbrido basado en las metodologías Adaptive Neuro Fuzzy Inferences System (ANFIS) y Probit.

Para comparar los resultados obtenidos por el modelo propuesto, este trabajo incluye la aplicación de dos técnicas de aprendizaje inteligente ampliamente utilizadas en la literatura; las Redes Neuronales (ANN), en particular el Perceptrón Multicapa (MLP) y Support Vector Machine (SVM).

Para cada modelo utilizado se realiza una optimización de sus parámetros más relevantes para la obtención de resultados. Para el caso de la red MLP se optimiza el número de neuronas de la capa escondida, la tasa de aprendizaje y el momentum. Para el modelo SVM, se obtienen los parámetros óptimos de costos  $C$  y de kernel  $\gamma$ . Por último, para el caso del modelo ANFIS y ANFIS-Probit se determinan los parámetros de funciones de membresía.

Los datos utilizados corresponden a la última versión de la Encuesta Financiera de Hogares 2014 elaborada por el Banco Central y cuyas observaciones utilizadas en esta investigación alcanzan los 3425 datos. El procesamiento de datos abarca el escalamiento y la aplicación de la técnica Smote para abordar el problema de desbalance de clases.

Los resultados obtenidos permiten indicar que el modelo propuesto es superior a los modelos benchmark en términos de Tasa de precisión y Tasa de precisión promedio además del Error tipo I.

**Palabras Clave: Hogares, Sobreendeudamiento, ANFIS, modelo híbrido, modelo predictivo**

---

## ABSTRACT

El aumento en los niveles de deuda de las familias de distintas partes del mundo ha atraído la atención de organizaciones locales y mundiales dedicadas a la prevención de riesgos financieros, y ha intensificado el interés por desarrollar métodos de detección temprana para el sobreendeudamiento en la población. El presente trabajo propone un modelo híbrido de Adaptive Neuro Fuzzy Inferences System (ANFIS) para la predicción del sobreendeudamiento de hogares, basado en una técnica estadística y Neuro Fuzzy. El modelo propuesto se comparó con los modelos Probit, Perceptrón multicapa (MLP) y Support Vector Machine (SVM). Se optimizan los parámetros más relevantes para el desempeño de cada técnica, y manejamos los problemas de equilibrio de datos a través de la técnica de sobre muestreo Smote. Utilizamos datos obtenidos de la Encuesta Financiera de Hogares del Banco Central (EFH) 2014 de Chile. Los resultados muestran que el modelo propuesto tiene un rendimiento significativamente mejor que los modelos de referencia en términos de la tasa de clasificación correcta, la tasa de clasificación correcta promedio y el error de tipo I. En consecuencia, este trabajo proporciona una innovadora comprensión del problema de sobreendeudamiento de hogares que puede ser de gran utilidad para distintas entidades gubernamentales enfocadas en prevenir el endeudamiento excesivo y mantener la estabilidad financiera.

The increase in debt levels of families in different parts of the world has attracted the attention of local and global organizations dedicated to the prevention of financial risks, and has intensified the interest in developing early detection methods for over-indebtedness in the population. The present work proposes a hybrid model of Adaptive Neuro Fuzzy Inferences System (ANFIS) for the prediction of household over indebtedness, based on a statistical technique and Neuro Fuzzy. The proposed model was compared with the Probit, Multilayer Perceptron (MLP) and Support Vector Machine (SVM) models. The most relevant parameters for the performance of each technique are optimized, and we manage the data balance problems through the Smote oversampling technique. We use data obtained from the Financial Household Survey of the Central Bank (EFH) 2014 of Chile. The results show that the proposed model has a significantly better performance

---

than the reference models in terms of the correct classification rate, the average correct classification rate and the type I error. Consequently, this work provides an innovative understanding of the problem of over-indebtedness of households that can be very useful for different governmental entities focused on preventing excessive indebtedness and maintaining financial stability.

**Keywords: Hogares, Sobreendeudamiento, ANFIS, modelo híbrido, modelo predictivo**

# ÍNDICE DE CONTENIDOS

<b>1. INTRODUCCIÓN</b>	<b>1</b>
1.1. MOTIVACIÓN	1
1.2. OBJETIVO PRINCIPAL	4
1.3. OBJETIVOS SECUNDARIOS	4
1.4. ALCANCE	5
<b>2. MARCO TEÓRICO</b>	<b>7</b>
2.1. Problemática	7
2.1.1. Descripción de la problemática	7
2.1.2. Propuesta de tema	9
2.2. Inclusión Financiera	10
2.2.1. Contexto mundial	11
2.2.2. Inclusión financiera en Chile	12
2.3. Endeudamiento	15
2.3.1. Concepto de Sobreendeudamiento	16
2.3.2. Endeudamiento en Chile	18
2.3.3. Perfil de los hogares endeudados de Chile	20
2.3.4. Estudios de análisis empírico	21
2.4. Algoritmos de aprendizaje inteligente	25
2.4.1. Modelos de datos desbalanceados	27
2.4.1.1. Synthetic Minority Over-sampling Technique	28
<b>3. METODOLOGÍA</b>	<b>30</b>
3.1. Modelos propuestos	30
3.1.1. Modelo Probit	30
3.1.2. Red neuronal	33
3.1.2.1. Estructura de una Red Neuronal Artificial	33
3.1.3. Multilayer Perceptron	34
3.1.4. Support Vector Machine	36
3.1.5. Adaptive Neuro Fuzzy Inference System	37
3.1.6. Modelo Híbrido	40
3.2. Data	42
3.2.1. Descripción de datos	42
3.2.2. Pre procesamiento de datos	43
3.2.2.1. Smote	44
3.2.2.2. Escalamiento de datos	44

<b>4. RESULTADOS</b>	<b>45</b>
4.1. Resultados	45
4.1.1. Resultados modelo Probit	45
4.1.2. Resultados modelo MLP	46
4.1.3. Resultados modelo SVM	48
4.1.4. Resultados modelo ANFIS	49
4.1.5. Resultados modelo Híbrido	50
4.2. Discusión de resultados	51
4.2.1. Wilcoxon test	51
4.2.2. Errores de clasificación	52
<b>5. ANEXOS</b>	<b>54</b>
5.1. Descripción de variables	54
5.2. Indicadores de endeudamiento	56
<b>Bibliografía</b>	<b>58</b>

# ÍNDICE DE TABLAS

2.1. Mediana de los indicadores de endeudamiento (pesos, porcentaje, número de veces). Fuente: Informe de Endeudamiento 2017, SBIF. . . . .	19
3.1. Estadísticas descriptivas. . . . .	42
3.2. Selección de variables. . . . .	43
3.3. Distribución de la variable independiente (%). . . . .	44
4.1. Resultados modelo de sobreendeudamiento con Probit. . . . .	45
4.2. Variación en la clasificación y promedio de clasificación según el número de nodos ocultos (%). . . . .	46
4.3. Variación en la clasificación promedio según la tasa de aprendizaje y momentum (%). . . . .	47
4.4. Resultados modelo de sobreendeudamiento con MLP. . . . .	48
4.5. Resultados modelo de sobreendeudamiento con SVM. . . . .	48
4.6. Resultados del modelo ANFIS con distintas funciones de membresía (%). . . . .	49
4.7. Resultados modelo de sobreendeudamiento con ANFIS. . . . .	50
4.8. Resultados del modelo híbrido Probit-ANFIS con distintas funciones de membresía (%). . . . .	50
4.9. Resultados modelo de sobreendeudamiento con Probit-ANFIS. . . . .	51
4.10. Tasa de precisión y Tasa de precisión promedio de los modelos construidos (%). . . . .	52
4.11. Wilcoxon signed-rank test para comparación de los resultados de clasificación obtenidos. . . . .	52
4.12. Error Tipo II y Tipo II de los modelos construidos (%). . . . .	53

# ÍNDICE DE FIGURAS

2.1. Número de cuentas transaccionales (millones de unidades). Fuente: Asociación de Bancos en base a SBIF . . . . .	13
2.2. Características socio-demográficas de personas bancarizadas y no bancarizadas en Chile (Porcentaje de las personas en cada grupo). Fuente: Asociación de Bancos en base a Banco Mundial, Global Findex 2014. . . . .	14
2.3. Endeudamiento de Hogares en Chile. Fuente: Elaboración propia con datos del Banco Central . . . . .	20
2.4. Hogares con deuda ( % del total de hogares en cada categoría de ingresos.)	21
2.5. Funcionamiento SMOTE . . . . .	28
3.1. Arquitectura Perceptrón Multicapa . . . . .	34
3.2. Support Vector Machine . . . . .	36
3.3. Arquitectura ANFIS . . . . .	38
3.4. Proceso de elaboración del modelo híbrido propuesto . . . . .	41

# CAPÍTULO 1

## INTRODUCCIÓN

### 1.1. MOTIVACIÓN

Desarrollo social y la estabilidad financiera son dos de los desafíos que deben enfrentar las sociedades modernas. Gran parte del mundo integra en sus programas de gobierno políticas, planes, y programas orientados a construir una sociedad más justa para sus habitantes, poniendo en primera línea la lucha contra de pobreza y la desigualdad de oportunidades.

Financieramente, uno de los principales enemigos de las sociedades actuales es el sobreendeudamiento, especialmente de sus hogares. Anderloni y Vandone [1] plantean la inexistencia de una definición única en la literatura para el fenómeno de sobreendeudamiento pero sostienen que su naturaleza es multidisciplinaria. Ntsalaze y Ikhide [2] plantean que el sobreendeudamiento representa, sin lugar a dudas, un riesgo importante para cualquier sistema financiero y la estabilidad del mismo, reafirman la incomprensión de su origen y mencionan que existen múltiples dificultades en su medición.

Analizar periódicamente la realidad social es quizás uno de los mecanismos más efectivos al momento de identificar las necesidades y riesgos de la población, principalmente de aquellas personas o grupos vulnerables. Detectar temprana y oportunamente situaciones que representen un riesgo de endeudamiento permite dar luces de lo que podría ser causa de su origen, como también aportar los antecedentes necesarios para la toma de decisiones en el diseño y aplicación de políticas públicas y planes sociales en materia de protección

financiera.

Trabajos como las de Betti, Dourmashkin, Rossi, y Ping Yin [3], cuyo enfoque es la cuantificación y caracterización del sobreendeudamiento en los países de la Unión Europea, y D'Alessio y Iezzi [4], quienes buscan definir y medir el sobreendeudamiento en los hogares con datos de la población Italiana, son solo algunas de las múltiples investigaciones que han realizado aportes significativos al conocimiento teórico del fenómeno de sobreendeudamiento en la población, ya que principalmente han permitido identificar y analizar factores que resultan determinantes para la existencia del sobreendeudamiento y por sobre todo han logrado avanzar en la caracterización de los afectados.

Gran parte de los trabajo han utilizado la econometría para justificar sus conclusiones, a través de modelos como análisis discriminante, regresiones lineales y logísticas, entre otros. Sin embargo, gracias a una amplia revisión literaria, es posible afirmar que gran parte las investigaciones realizadas a la fecha presentan una importante limitación que es la capacidad de aplicar los modelos desarrollados para lograr predecir la condición en base a nuevos datos. Esta limitación encontrada representa un vacío en el campo del conocimiento respecto al sobreendeudamiento que no ha sido abordado por nuevos investigaciones y que es fuente de inspiración para el desarrollo de este trabajo.

La econometría tradicional ha permitido que las investigaciones realizadas logran su objetivo prioritario de analizar separadamente la influencia de una u otra variable o factor, pero no abordó eficientemente la facultad de predecir. Anticiparse o predecir los fenómenos a partir de información existente es una de las características más buscadas y anheladas por los estudios modernos.

De acuerdo con la situación financiera y económica actual, muchos países han puesto sus esfuerzos en perfeccionar sus sistemas de protección social en endeudamiento y pobreza, así como en múltiples otros aspectos de la vida diaria. El objetivo es generar nuevos y mejores mecanismos e instrumentos metodológicos que permitan un estudio de mayor calidad y que cubra distintos aspectos del fenómeno en observación.

Bajo esta perspectiva, la aplicación de técnicas derivadas de la Inteligencia Artificial surge como una innovadora alternativa a la econometría tradicional, que ha permitido a los investigadores avanzar y mejorar el conocimiento que se tiene, ya que han demostrado

un dominio importante en la predicción de problemas reales. Dentro de las principales técnicas computacionales que han permitido mejorar el campo del conocimiento en el área financiera y económica destacan las redes neuronales (ANN) (para ejemplos y aplicaciones ver: [5]), árbol de decisión (DT), support vector machine (SVM), entre otros (ver [6]). A pesar de la innegable utilidad de estas nuevas metodologías, su aplicación, en el ámbito financiero, ha sido extensamente desarrollada en temas relativos al interés macroeconómico y de negocios pero sus aportes en materia de hogares y su comportamiento es escasa.

La combinación de metodologías clásicas con técnicas computacionales o entre algoritmos inteligentes ha adquirido especial atención puesto que surgen como el siguiente escalón en la búsqueda de mejores modelos aplicados a problemas reales. La combinación de técnicas generalmente conocida como modelos híbridos o ensamblados propone extraer el máximo beneficio que cada técnica ofrece y combinarlos logrando superar los resultados que se obtendrían con cada técnica por separado.

Inspirada en los resultados obtenidos por trabajos basados en técnicas de la inteligencia artificial relativos temas como estrés financiero y predicción de la bancarrota, este trabajo busca aplicar similares procedimientos pero con énfasis en una unidad económica más pequeña: los hogares o familias.

El objetivo de este trabajo es proponer una metodología innovadora basada en la inteligencia artificial a través de un modelo híbrido que busca dar una nueva perspectiva al estudio del sobreendeudamiento. La principal motivación es aportar una potencial opción de análisis a un problema de la vida real y de esta forma reducir los riesgos asociados para quienes pueden ser afectados.

Con este trabajo se entrega un procedimiento que involucra el modelamiento del comportamiento de los hogares con sobreendeudamiento como paso inicial y que se enfoca principalmente en entrega una predicción de esta condición sobre nuevos datos con un grado de asertividad elevado.

## 1.2. OBJETIVO PRINCIPAL

El objetivo principal de este estudio es proponer una nueva metodología para mejorar la clasificación y predicción de la condición de sobreendeudamiento de hogares mediante el uso de un modelo híbrido basado en Adaptive Neuro Fuzzy Inference System en comparación al uso de un modelo econométrico Probit y las metodologías inteligentes Multilayer Perceptron y Support Vector Machine.

## 1.3. OBJETIVOS SECUNDARIOS

- Aportar a la literatura existente en relación a modelos enfocados en predicción del sobreendeudamiento de hogares mediante los resultados de las nuevas metodologías desarrolladas.
- Utilizar los datos brutos obtenidos desde el Banco Central de Chile correspondientes a la Encuesta Financiera de Hogares 2014 para el diseño de un modelo de sobreendeudamiento de hogares.
- Optimizar los parámetros relevantes para las técnicas Multilayer Perceptron (MLP), Support Vector Machine (SVM) y Adaptive Neuro Fuzzy Inference System (ANFIS) utilizados como benchmark.
- Procesar los datos seleccionados con la técnica Adaptive Neuro Fuzzy Inference System (ANFIS) en combinación con el modelo econométrico Probit.
- Validar los resultados obtenidos por el modelamiento mediante la predicción de nuevos datos disponibles.
- Concluir la superioridad del modelo propuesto a través del test de Wilcoxon disponible en la literatura.

## 1.4. ALCANCE

La innovación de la metodología híbrida propuesta en este trabajo se enfoca principalmente en el uso del modelo Adaptive Neuro Fuzzy Inference System como base para la predicción del sobreendeudamiento de hogares. El análisis basa su procedimiento y conclusiones utilizando los datos de la Encuesta Financiera de Hogares 2014 del Banco Central de Chile.

A pesar de que existen investigaciones con base en estos datos y de índole financiera en la literatura chilena y del mundo, estos trabajos dirigen su interés en el análisis teórico de factores que condicionan el endeudamiento en personas u hogares, y para ello hacen uso de la econometría tradicional (como por ejemplo regresiones lineales o logísticas), pero carecen de interés por el área de predicción para la condición misma, lo que se traduce a un desafío al momento de buscar métodos y/o procedimientos para abordar este vacío en la literatura.

El alcance de esta propuesta logra abordar dos aspectos; modelamiento y predicción, que usualmente no se han complementado bajo un mismo procedimiento. Es importante considerar que un punto fundamental y esencial de un buen modelo de sobreendeudamiento es su capacidad de identificar correctamente aquellos individuos en riesgo, especialmente en individuos que se encuentran fuera de la muestra (i.e predecir la condición).

De forma compacta, este trabajo se considera innovador por:

- Proponer un modelo híbrido basado en Adaptive Neuro Fuzzy Inferences System (ANFIS) en conjunto con la técnica clásica Probit que a nuestro conocimiento no ha sido utilizada en otros estudios de la misma índole.
- Aplicar una técnica computacional al sobreendeudamiento de hogares que es un concepto frecuentemente analizado desde la perspectiva de la teoría y cuya investigación relacionada a la predicción no ha sido ampliamente desarrollada.
- Enfrentar el problema de datos desbalanceados a través del método Smote logrando beneficiar la capacidad predicativa del modelo.
- La propuesta apunta a ofrecer una metodología real y completa que puede ser

considerada y de gran utilidad por las distintas entidades Gubernamentales en sus políticas y planes de prevención, para apoyar las decisiones en materia de desarrollo y estabilidad de sus economías locales.



## CAPÍTULO 2

# MARCO TEÓRICO

### 2.1. Problemática

#### 2.1.1. Descripción de la problemática

Durante los últimos años Chile ha experimentado una gran expansión en el acceso al crédito por parte de los hogares en conjunto, como también, por cada uno de sus habitantes. Estos avances han tenido implicancias positivas en términos de profundización e inclusión financiera, ofreciendo más alternativas de pago y financiamiento a los consumidores. Sin embargo, dichos elementos también se han convertido en mayores riesgos financieros para deudores y acreedores. En términos concretos, los niveles de endeudamiento de la población y en particular, de las familias en Chile han aumentado desde un 36 % del PIB en 2010 a un 46 % del PIB en 2017 [7].

El nivel de endeudamiento de los hogares y su crecimiento sostenido durante las últimas décadas, ha comenzando a despertar inquietud en nuestro país, y especialmente de sus actores principales; las familias. Junto con constituir una oportunidad social para muchas de ellas, el endeudamiento se ha transformado también en un factor de alto riesgo, principalmente por la facilidad de los créditos de consumo otorgados por casas comerciales, cuyas tasas de interés son considerablemente altas, introduciendo a las familias en un círculo de deuda del cual resulta cada vez más complejo salir y que puede finalmente aumentar la vulnerabilidad de un hogar.

Los datos entregados por el Fondo Monetario Internacional (FMI) señalan a Chile como el país con mayores niveles de deuda de hogares de América Latina respecto al PIB. Según el organismo, Chile supera a Panamá, Costa Rica y Brasil, quienes han registrado un 38 %, 35 % y 27 % del PIB respectivamente. Muy por debajo aparecen Uruguay (10 %), Paraguay (7 %) y Argentina (4 %). De acuerdo a especialistas, estas cifras obedecen a los mayores niveles de bancarización del país y también a las amplias opciones que existen para conseguir créditos hipotecarios a largo plazo.

La Encuesta de Ocupación y Desocupación en el Gran Santiago de la Universidad de Chile señala que, el 58.1 % de los hogares reporta tener algún tipo de deuda. Por otra parte, el promedio de la deuda de los hogares ha alcanzado un máximo de 71.1 % del ingreso disponible el 2017 superando así los registros del año 2016 en aproximadamente 3.7 puntos porcentuales lo que significa un alarmante nuevo record. En detalle, las cifras oficiales entregadas por el Banco Central señalan que la constitución de la deuda de los hogares se basa principalmente en créditos, tanto bancarios; para la vivienda (37.4 %), como de consumo (18.2 %) [8].

Los datos señalados dan cuenta de una preocupante realidad de vulnerabilidad y estrés financiero en un inmenso número de hogares, que únicamente entrega cifras a partir de los ingresos. Si los recursos obtenidos mensualmente se destinan a pagar, en gran medida, el consumo efectuado en el pasado, y los hogares no pueden hacer frente a sus necesidades actuales, la pobreza real puede ser mayor a la que indican las cifras oficiales, al dejar al margen este aspecto, no estaría respondiendo a la realidad. Por esta razón, se hace necesaria la incorporación de la variable endeudamiento en la medición de la pobreza, lo que incentivaría, además, la preocupación social y política por una cuestión creciente en el país, que posiblemente tendrá consecuencias importantes en la lucha por la erradicación de la pobreza.

La relación del sobreendeudamiento con la pobreza no puede ser eludida, y es susceptible de ser estudiada con análisis empíricos que permitan verificar de manera sólida esta relación. La toma de conciencia respecto de este tema y la medición de su impacto permitirán buscar soluciones desde la política pública, que lleven a la autoridad y a todos los actores del sistema a fomentar un endeudamiento sostenible, que no ponga en riesgo el

presente y el futuro de los hogares más vulnerables.

El objetivo este trabajo es proponer nuevas metodologías de análisis para la condición de sobreendeudamiento de hogares para así ahondar en la protección social a las familias chilenas.

### **2.1.2. Propuesta de tema**

El estudio del endeudamiento y de la predicción de incumplimiento de las obligaciones financieras por parte de los hogares permite identificar las vulnerabilidades de los mismos frente a choques macroeconómicos o de otro tipo, que puedan afectar la disponibilidad de recursos destinados al pago de deudas contraídas con el sistema financiero. De esta forma es posible establecer los efectos sobre la estabilidad financiera asociados al deterioro económico dentro de los hogares y promover las medidas necesarias para minimizar su impacto.

Frente al evidente impacto negativo que puede generar el desarrollo de un sobreendeudamiento en la población, controlar esta condición en un número significativo de hogares es un nuevo desafío social que aún no ha sido valorado en toda su importancia. Sin embargo, si la superación de la pobreza es realmente una preocupación nacional, es necesario promover un cambio en los patrones de endeudamiento que verdaderamente la haga posible, se deben realizar todos los esfuerzos para entender y analizar de forma óptima la condición en sí misma y aquellos factores determinantes para su desarrollo. Un punto de partida es la existencia de una capacidad predictiva que facilite la mayor fiscalización de las instituciones financieras encargadas de proteger financieramente a los hogares en riesgo.

Bajo la óptica de la estabilidad financiera, este estudio plantea una metodología orientada a facilitar esta comprensión de la deuda, permitiendo identificar de forma eficiente los hogares de una muestra que presentan una condición de sobreendeudamiento, además de proponer un modelo con capacidad predictiva significativamente superior a los estudios desarrollados a la fecha en materia de sobreendeudamiento de hogares.

## 2.2. Inclusión Financiera

La inclusión financiera es un concepto amplio que no presenta un consenso general sobre su definición, pero en cada una de las existentes se menciona como factor determinante para el bienestar y desarrollo financiero de un país, que involucra tanto el acceso a productos y servicios financieros, como el hábito de usarlos y los conocimientos adecuados para desenvolverse como usuarios del sistema financiero, transformándose así en una condición fundamental para la vida cotidiana de las personas.

La inclusión financiera adquiere gran importancia en las últimas décadas convirtiéndose en un indicador de desarrollo financiero y económico sobre el cual, distintas naciones y sus políticas públicas apuntan a mejorar. A nivel internacional existe la Alianza para la Inclusión Financiera implementada por el G-20, cuyo objetivo es hacer que los servicios financieros sean más accesibles para las personas no bancarizadas alrededor del mundo. Esta alianza incluye más de 90 países que trabajan en conjunto para ayudar a los países a diseñar e implementar soluciones de alto impacto para satisfacer sus necesidades y desafíos en materia de inclusión financiera.

En el 2011, el Banco Mundial puso en marcha Global Findex, la base de datos más completa del mundo sobre la inclusión financiera. Este instrumento ofrece un panorama riguroso y multidimensional con información detallada sobre la manera en que los adultos en más de 140 economías acceden a las cuentas, realizan pagos, ahorran, piden préstamos y manejan el riesgo. A la encuesta inicial le siguieron una segunda en el 2014 y una tercera en el 2017.

Debido a su importancia, los beneficios potenciales de la inclusión financiera al desarrollo de las economías modernas, están ampliamente documentados en la literatura internacional. Se ha señalado que la inclusión financiera ayuda a suavizar ingreso y consumo de la población [9] propicia la eficiencia en la asignación de recursos [10], potencia el crecimiento económico (Cecchetti, et al., 2006) y ayuda a disminuir la pobreza y la desigualdad [11].

Guerrero, Espinosa y Focke [12] confirman que un sistema financiero bien desarrollado es prerequisite para el crecimiento económico. Además declaran que el fortalecimiento

y desarrollo de los sectores de menores ingresos, especialmente de las economías rurales ya que un mayor acceso a servicios bancarios dinamiza las actividades de las pequeñas economías, genera capacidades para aprovechar de forma óptima los recursos, eleva los niveles de ingresos y disminuye las imperfecciones de los mercados contribuyendo a activar la economía a nivel local y nacional.

Anand y Prasad [13] indican que la inclusión financiera mejora en promedio la suavización del consumo, lo que tiene impacto en las decisiones de política. Dabla-Norris y Srivisal (2013) muestran que una mayor profundidad financiera disminuye la volatilidad del crecimiento y del consumo, lo cual ceteris paribus se traduce en un mayor bienestar a largo plazo para la sociedad.

### 2.2.1. Contexto mundial

De acuerdo con la última encuesta realizada el año 2017 por el Banco Mundial, los datos de Global Findex revelan que globalmente el 69 % de la población adulta (mayor o igual a 15 años)- 3800 millones de personas- posee una cuenta en una institución financiera, un aumento de 18 y 7 puntos porcentuales desde el 2011 y 2014 respectivamente [14]. Esto se traduce en que 515 millones de adultos alrededor del mundo abrieron una cuenta en una institución financiera o a través de un proveedor de dinero móvil.

Algunas economías registran un aumento considerable en el número de titulares de cuentas, lo que difiere de otras economías que han registrado progresos más lentos, condicionados en su mayoría por grandes disparidades entre hombres y mujeres y entre ricos y pobres. Es así como, en economías de altos ingresos, el 94 % de los adultos tienen una cuenta; en economías en desarrollo, el 63 %.

Gran parte del progreso reciente se vio impulsado por los pagos digitales, las políticas gubernamentales y una nueva generación de servicios financieros a los que se accede a través de teléfonos celulares e internet. En detalle, se han registrado aumentos significativos en el uso de teléfonos celulares y de Internet para realizar operaciones financieras. Esto ha contribuido a que la proporción de titulares de cuentas que envían o reciben pagos a través de medios digitales aumente del 67 % al 76 % a nivel mundial, y del 57 % al 70 % en el mundo en desarrollo.

Pese al crecimiento en las tasas de inclusión financiera se estima que cerca de 1700 millones de adultos alrededor del mundo permanecen no bancarizados, es decir, aún no poseen una cuenta en una institución financiera o a través de un proveedor de dinero móvil. Un aspecto importante de esta cifra es que, prácticamente todos estos adultos no bancarizados viven en países en desarrollo (por ejemplo India, México, Nigeria, Pakistán, entre otros).

Un porcentaje no despreciable de los adultos no bancarizados pertenecen a los sectores económicos más pobres de sus países. Se estima que la mitad de los adultos no bancarizados pertenecen al 40 % de los hogares más pobres dentro de su economía, el 60 % más rico representa la mitad restante.

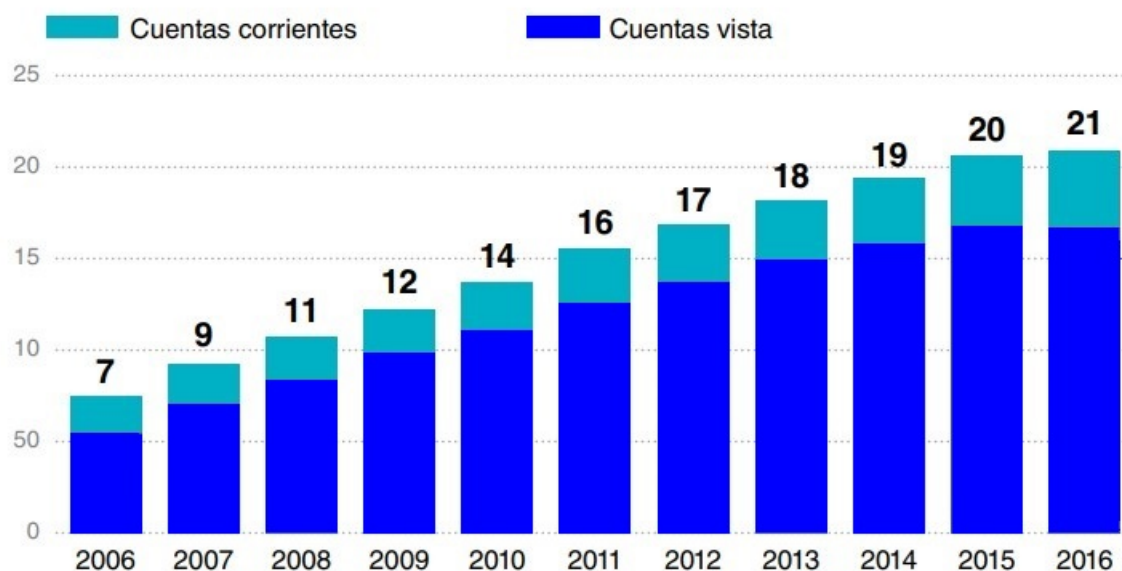
### **2.2.2. Inclusión financiera en Chile**

Cercano a 1970, del sistema bancario chileno prestaba servicios a un segmento reducido de la población. Su atención se centraba en las empresas de mayor tamaño y a los hogares de ingresos más altos. Esta segregación provocaba que muchas empresas y personas quedaran fuera del sistema existente hasta ese momento. Esta situación se ha modificado en el curso de las últimas décadas, al igual que en gran parte del mundo. La industria financiera ha adoptado la inclusión financiera puesto que representa una oportunidad de negocios, porque supone la ampliación de su base de clientes y de sus volúmenes de actividad, es por esta razón que los bancos han desplegado esfuerzos para llegar a nuevos segmentos de la población. Ello se ha visto facilitado por el propio desarrollo de la economía y el consiguiente aumento del ingreso de las personas. Lo cierto es que los avances en materia de acceso a los servicios han sido enormes, especialmente si se los evalúa en perspectiva.

Para el año 2016, la Superintendencia de Banco e Instituciones Financieras (SBIF) presentó el primer Informe de Inclusión Financiera en Chile, donde el principal objetivo fue contribuir al diagnóstico de la situación de inclusión financiera en el país, mejorar la información disponible en esta materia y sentar un punto de referencia para futuras mediciones. De acuerdo con el Informe y dentro de sus principales resultados se destacó que, para el año 2016, el 98 % de la población adulta (mayor o igual a 15 años) contaba con al menos un producto financiero (activo o pasivo) y un 63 % de los adultos eran titulares

de una cuenta en una institución financiera. Dentro de los productos más se encontraban aquellos asociados a la administración del efectivo como son las cuentas corrientes y cuentas vista (con un 76 % de cobertura, como porcentaje de la población adulta) y el ahorro (69 %).

Respecto a las cuentas transaccionales (cuentas corrientes y cuentas vista), la evolución de Chile destaca positivamente, ya que la cantidad de cuentas se ha triplicado en la última década. La Fig.2.1. muestra el crecimiento de las unidades desde el año 2006.



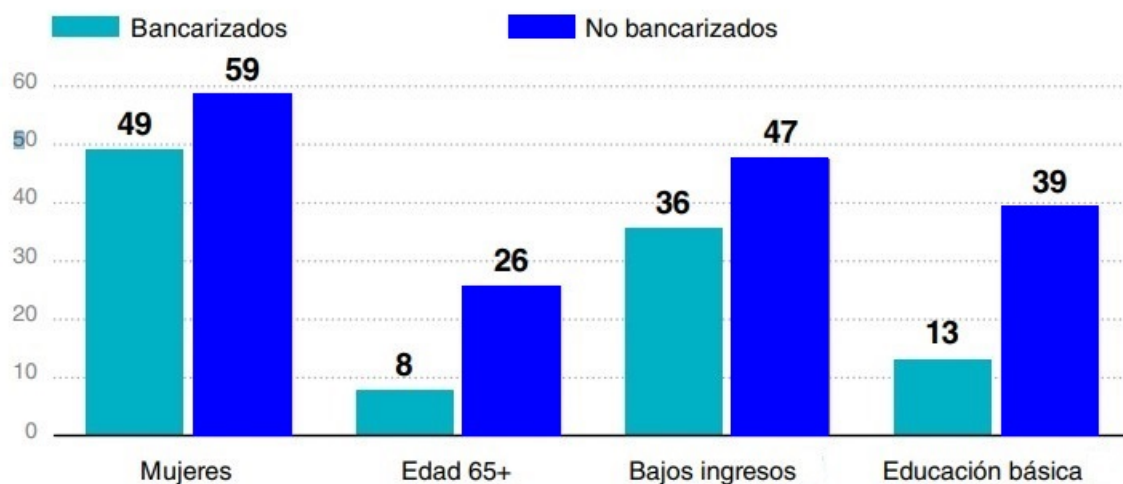
**Figura 2.1:** Número de cuentas transaccionales (millones de unidades). Fuente: Asociación de Bancos en base a SBIF

Relativo al acceso a específico de cada producto financiero, los indicadores del sistema bancario chileno indican que, un tercio de la población adulta chilena posee alguna tarjeta de crédito bancaria y sólo 15 % posee una cuenta corriente. Asimismo, una de cada tres personas posee una tarjeta de crédito no bancaria. El 18 % posee un crédito en cuotas y un 7 % un crédito para la vivienda. Sólo el 4 % de la población posee un depósito a plazo.

Esta considerable expansión en los niveles de inclusión y acceso a distintos productos y servicios financiero ha permitido que gran parte de la población mejore su calidad de vida. Uno de los elementos más importantes en este aumento del acceso a crédito corresponde a la expansión del sector no bancario, como el crédito que entregan las casas comerciales. En efecto, estas han aumentado el acceso al crédito para un sector importante de la población

típicamente restringido [15], permitiendo que las personas puedan absorber descalces temporales entre ingresos y gastos e incrementar por esta vía su bienestar.

A pesar del crecimiento sostenido en la inclusión y acceso financiero, similar a lo acontecido internacionalmente, en Chile existe un grupo de personas que siguen siendo no bancarizados. De acuerdo con la SBIF, este grupo se caracteriza por una participación elevada de mujeres y de personas de tercera edad. Como lo muestra la Fig.2.2., el 59 % de los no bancarizados son mujeres, comparado con 49 % entre los que tienen cuentas, y las personas mayores de 65 años equivalen a 26 % de los no bancarizados. Asimismo, las personas sin cuenta bancaria tienen mayor probabilidad de tener un bajo nivel educacional<sup>1</sup> y bajos ingresos<sup>2</sup>. El 39 % de los no bancarizados tiene educación media incompleta o menos, comparado con sólo 13 % entre los bancarizados, y casi la mitad de las personas sin cuenta pertenece a hogares de menores ingresos.



**Figura 2.2:** Características socio-demográficas de personas bancarizadas y no bancarizadas en Chile (Porcentaje de las personas en cada grupo). Fuente: Asociación de Bancos en base a Banco Mundial, Global Findex 2014.

Según Dupas et, al.[16], la no bancarización chilena sería predominantemente voluntaria (falta de demanda) y no necesariamente asociada a dificultades de acceso o asequibilidad (restricciones de oferta). En su estudio los autores comparan a Chile con dos economías de menor nivel de desarrollo; Malawi y Uganda, y evalúan el efecto de remover las barreras de

<sup>1</sup>Personas con educación básica completa o menos (pero sin educación media completa)

<sup>2</sup>Personas que pertenecen al 40 % de los hogares de menores ingresos

acceso a cuentas (costos y procesos legales) para hogares no bancarizados. Sus resultados obtenidos concluyen que sólo un 17 % de los hogares chilenos accedió a la oferta, mientras que en los países en comparación, la cifra superó el 50 %, concluyendo de esta manera, que los hogares en Chile optan voluntariamente por seguir sin bancarización.

La encuesta del Banco Mundial (2014) muestra resultados similares al estudio anterior, ya que el 76 % de las personas no bancarizadas alude razones que indican ausencia de necesidad o interés (falta de demanda). Entre las otras razones identificadas en esta encuesta, el 44 % de los encuestados menciona el costo, lo que si bien contrasta con el hecho de que existan cuentas a la vista de bajo costo y cuentas de ahorro libres de cobros por mantención, es un aspecto a considerar en el diseño de futuras estrategias de inclusión específicas a este segmento.

Sin embargo, simultáneo al mejoramiento del acceso financiero, se ha producido un preocupante efecto secundario; el endeudamiento de la población. Elevados niveles de endeudamiento pueden afectar la capacidad de los hogares para cumplir con sus compromisos financieros y hacerlos más vulnerables a shocks agregados, generando efectos negativos sobre la estabilidad financiera. Es por esta razón, entre otras, que el monitoreo del endeudamiento de las personas es un tema de especial relevancia para diversas entidades de fiscalización financiera.

## 2.3. Endeudamiento

Un mayor y mejor acceso a los servicios financieros desarrolla un efecto secundario que es el aumento del endeudamiento por parte de la población. Varias teorías económicas buscan explicar el endeudamiento ya sea a través teoría del ingreso permanente y ciclo de vida, los hábitos de gasto y ahorro de las personas, aspiraciones de estilos de vida, entre otros [17]. Si bien el uso del crédito ha permitido a que las familias puedan gastar en un consumo presente, comprometiendo sus ingresos futuros, y de esta manera suavizar la senda de consumo a través de las diferentes etapas de la vida, las mejoras en bienestar asociadas al uso del crédito pueden verse cuestionadas cuando el nivel de deuda es excesivo y se produce una carga financiera que los hogares no pueden manejar de manera adecuada,

clasificando como un hogar sobre endeudado.

El fenómeno del sobreendeudamiento de hogares no es nuevo para la economía, ya que el riesgo de su aparición está, en cierta medida, vinculado al desarrollo y evolución del mercado financiero en sí mismo [18]. Sin embargo, la condición de un endeudamiento excesivo de hogares se vuelve perjudicial cuando el sector financiero tiene pocas regulaciones, la gente tiene menos dificultades para endeudarse y la probabilidad de no poder cumplir con los compromisos adquiridos es alta [19].

Un crecimiento desproporcionado del nivel de deuda de los hogares puede acarrear consecuencias importantes como se mencionó anteriormente. Los hogares más endeudados, que son los que sufren mayores cargas financieras, disfrutan de menor capacidad para hacer frente a cambios no esperados y, en consecuencia, sufren una mayor probabilidad de caer en situación de impago. La imposibilidad de cumplir las obligaciones financieras adquiridas en una cantidad significativa de hogares implicaría. Evidentemente, efectos perjudiciales en todo el sistema financiero que los ampara.

### 2.3.1. Concepto de Sobreendeudamiento

La conceptualización del sobreendeudamiento de los hogares es variada y a menudo confusa. El debate internacional se ha referido muchas veces al sobreendeudamiento en una forma limitada, enfocada en la economía y el individuo, estableciendo diferencias según el punto de vista bajo el que estudia la condición e ignorando la dinámica inherente de un fenómeno económico. En su análisis, la deuda excesiva debiese ser entendida como indicador de tendencias socioeconómicas y políticas mucha más amplia.

De forma intuitiva, participantes del sistema financiero convienen en que el sobreendeudamiento sucede una vez que se ha incurrido en deuda excesiva. Esta aproximación deja en el aire definiciones secundarias como ¿Qué constituye una deuda “excesiva” y cómo se define el umbral?, ¿Cómo medir cuantitativamente la gravedad de esta condición? ¿Qué indicadores son adecuados y cuál debería ser la unidad de medida? La definición por ende no es única y su interpretación y medición refleja objetivos específicos de disciplinas variadas [20].

Anderloni y Vandone (2008) describen el sobreendeudamiento como un fenómeno

multifacético con influencia en el ámbito económico [21], legal [22], psicológico [23] y social [24]. Autores han buscado definir el fenómeno de manera conceptual a través de las causas [25] y consecuencias del exceso de la deuda [26] además del rol de la educación financiera [27]. También de manera empírica a través de modelos de la estadística clásica (i.e regresión lineal o logística, series de tiempo, análisis de corte transversal) ([28]; [29]; [30]) con énfasis como la vulnerabilidad financiera [31] y el bienestar general [32].

Junto con la comprensión conceptual del sobreendeudamiento, durante las últimas décadas, se ha desarrollado un interés por la cuantificación de esta condición. Desde una perspectiva macroeconómica, se han elaborado índices complejos destinados a captar los diferentes aspectos del fenómeno<sup>3</sup> o la subjetividad de los prestatarios [20]. La medida y la cuantificación han sido, desde luego, dos de las preocupaciones políticas más importantes de quienes deben tomar decisiones. Guérin et al. [34] indican que, si bien los indicadores financieros son útiles y son utilizados para cuantificar el costo de la deuda para la sociedad, al momento de entender las prácticas de endeudamiento, las motivaciones y razones es necesario examinar en detalle y por separado del conjunto global, los hogares afectados y cómo estos perciben y experimentan el endeudamiento.

En la actualidad son numerosos los bancos centrales de distintos países que han aumentado su preocupación en entender el comportamiento financiero de los hogares, así como la obtención de datos e información rigurosa y periódica de los mismos [35]. El Banco de la Reserva Federal de Nueva York señala una serie de razones para ello [36]. En primer lugar, la crisis financiera internacional de 2008 puso de manifiesto que el entendimiento y la anticipación del comportamiento de los hogares es esencial para lograr una economía fuerte y resistente. En segundo lugar, es primordial averiguar cómo afectan realmente las intervenciones dirigidas a cambiar los incentivos de los hogares el comportamiento de los mismos. Por último, las investigaciones sobre las finanzas de los hogares ayudarán a apoyar el compromiso con la comunidad de los bancos centrales y el desarrollo de su misión y deberes. A la hora de lograr un mayor entendimiento de las finanzas de los hogares, los bancos centrales destacan la importancia que tiene la elaboración de bases de datos con

---

<sup>3</sup>Sobre la base de iniciativas europeas, se elaboró un Índice de Vulnerabilidad Financiera del Consumidor en Sudáfrica [33]. Incluye cuatro sub indicadores: vulnerabilidad del ingreso, vulnerabilidad de los ahorros, vulnerabilidad del gasto y vulnerabilidad por servicio de la deuda.

información de los comportamientos financieros y económicos de los hogares. Cada vez un mayor número de bancos centrales está promoviendo, bien sea desde el propio banco central o con el apoyo de otras instituciones –como ministerios o institutos nacionales de estadística-, la elaboración de bases de datos del lado de la demanda del consumidor financiero. En ese sentido, el informe elaborado por el Banco de Inglaterra [37] menciona la necesidad del uso de nuevos datos, metodologías y enfoques que ayuden a entender el comportamiento financiero y económico de los hogares.

En términos de estabilidad financiera, esta investigación argumenta que la comprensión del sobreendeudamiento mediante la identificación eficaz de los individuos bajo esta condición a través de una modelo de clasificación permite un mejor análisis de los factores influyentes y determinantes de aquellos hogares que están sobre endeudados. La comprensión del sobre endeudamiento como variable económica es fundamental para el funcionamiento y evolución estable de cualquier mercado financiero y sus actores principales.

### 2.3.2. Endeudamiento en Chile

Para el estudio del endeudamiento, entidades tales como la Superintendencia de Bancos e Instituciones Financieras (SBIF) y el Banco Central de Chile, han elaborado una serie de indicadores para monitorear la capacidad de tanto personas como de hogares, para hacer frente a sus obligaciones financieras y a la vez extraer información relevante de su evolución.

La SBIF, analiza el comportamiento de personas naturales con obligaciones de consumo o vivienda respecto a sus niveles de deuda. A través del Informe de Endeudamiento se definen los indicadores deuda, carga financiera y apalancamiento (definiciones en Anexo 5.2).

La versión más reciente del Informe de Endeudamiento publicada en 2017 [38] destaca que los indicadores de endeudamiento de los clientes bancarios aumentaron en el último año. Como muestra la Tabla 2.1, para el año 2017, el nivel de deuda representativo, definido a partir de la mediana de la distribución, fue de 2.6 millones de pesos, esta cifra aumentó en 7,3 % real en comparación a igual fecha del año anterior. Análogamente, los indicadores

de carga financiera y apalancamiento superaron a los observados doce meses antes.

Indicadores de endeudamiento			
	2015	2016	2017
Deuda	2.234.762	2.436.704	2.663.435
Carga Financiera	22.6	24.7	24.9
Apalancamiento	4.6	4.7	5

TABLA 2.1: Mediana de los indicadores de endeudamiento (pesos, porcentaje, número de veces). Fuente: Informe de Endeudamiento 2017, SBIF.

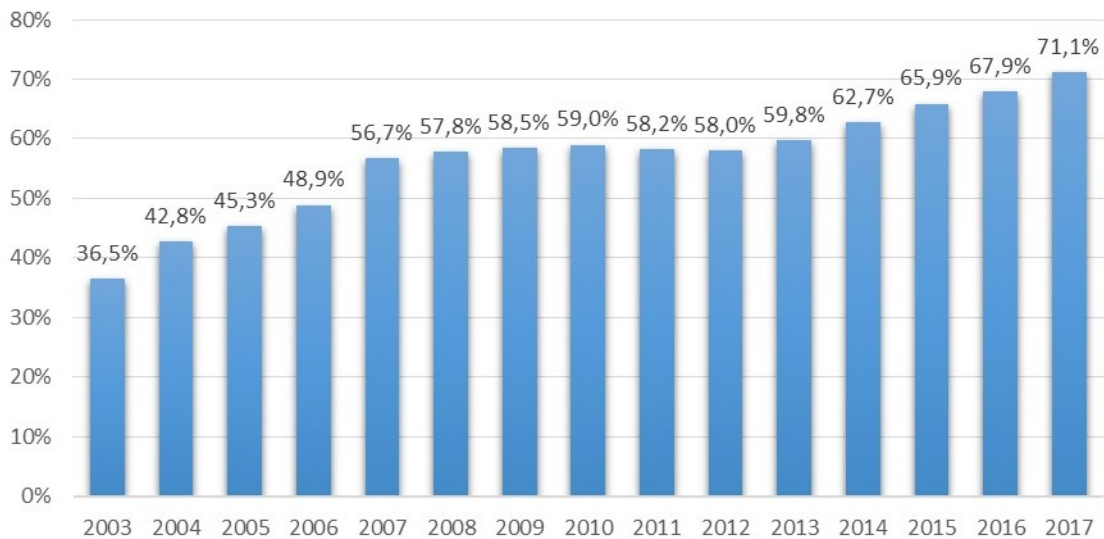
Por su parte, el Banco Central elabora semestralmente el Informe de Estabilidad Financiera (IEF), para analizar y dar a conocer los sucesos más recientes del ámbito macroeconómico y financiero que inciden en la estabilidad financiera de la economía nacional. Evolución del endeudamiento de los principales usuarios del crédito, el desempeño del mercado de capitales y la capacidad del sistema financiero son solo algunos de los temas de interés que se abordan en cada publicación. A través de este informe se entregan los principales indicadores financieros utilizados para en análisis temporal de comportamiento financiero de los hogares.

Dentro de los indicadores empleados se destaca la Razón Deuda a Ingreso (RDI) y la Razón Carga Financiera a Ingreso (RCI) (definiciones en Anexo 5.2). Estos indicadores son utilizados por un número de bancos centrales de países desarrollados para análisis tanto de estabilidad macroeconómica como de estabilidad financiera.

La versión del IEF 2018 [39] afirma que los hogares de Chile han aumentado el nivel de endeudamiento durante los últimos dos años. La información disponible ha permitido detectar que el endeudamiento de los hogares (RDI), entre los años 2001 y 2016 pasó de 35 a 65, mientras que la carga financiera (RCI) pasó de 9 a 15 (ambos casos respecto al porcentaje del ingreso disponible),

### 2.3.3. Perfil de los hogares endeudados de Chile

Las cifras de cuentas nacionales más recientes, publicadas por el Banco Central respecto al endeudamiento de hogares en Chile, indica que el stock<sup>4</sup> de la deuda alcanzó en 2017 un 71,1 % del ingreso disponible, superando así en un tres por ciento el cierre del año 2016. El aumento en los niveles de endeudamiento no es un evento sorpresivo, ya que tal como lo refleja la Figura 2.3, el crecimiento ha sido persistente durante los últimos años.



**Figura 2.3:** Endeudamiento de Hogares en Chile. Fuente: Elaboración propia con datos del Banco Central

Frente a esta situación, y gracias a la última versión de la Encuesta Financiera de Hogares (EFH), gestionada por el Banco Central durante el año 2014, las cifras más importantes relativas al endeudamiento de la población chilena han sido actualizadas y en consecuencia, han permitido identificar con mayor claridad el perfil de aquellos hogares con mayor riesgo financiero.

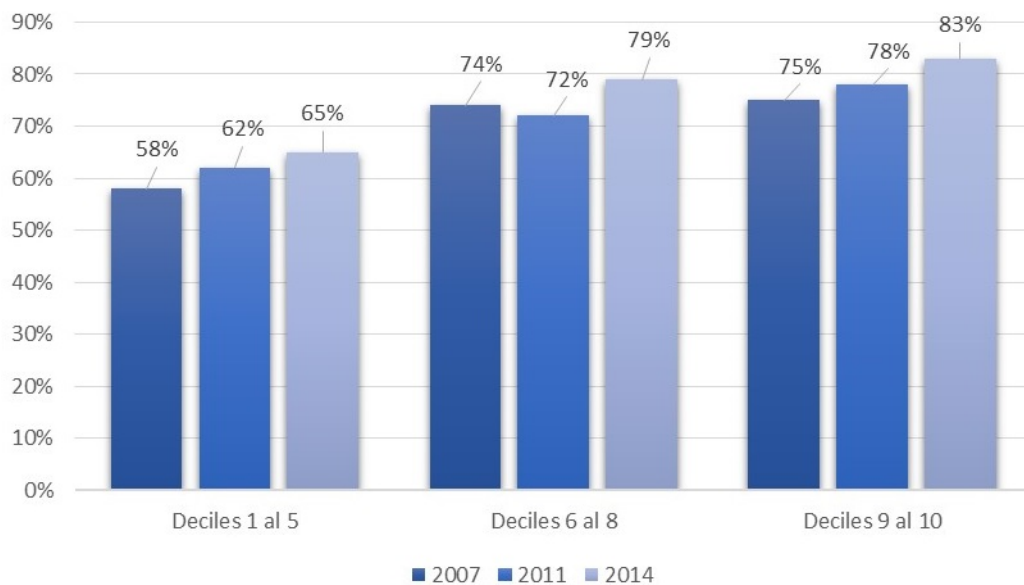
El Ministerio de Desarrollo Social en su Encuesta CASEN utiliza los deciles como una forma de clasificar a las familias según su nivel de ingresos. La asignación a uno u otro decil se determina según el cálculo del promedio de ingresos considerando la cantidad de personas que conforman ese hogar. De esta forma existe desde el decil 1, representa a la población con la condición socioeconómica más vulnerable, hasta el decil 10, que

<sup>4</sup>La relación deuda-ingreso corresponde al stock de pasivos mantenidos por los hogares al cierre del año, respecto del flujo total de los ingresos obtenidos en el año.

representa a las personas de mayores ingresos del país.

La EFH de 2014 permitió revelar que un 73 % de las familias chilenas registran obligaciones financieras. Respecto a la primera medición de esta encuesta, realizada el año 2007, esta cifra ha aumentado en siete puntos porcentuales. Sin embargo, se destacan principalmente dos aspectos importantes en este aumento; la cantidad y el tipo de deuda. Ambos items han desarrollado características diferentes según el estrato de ingreso de los hogares.

Un análisis en detalle del nivel de endeudamiento de los hogares, respecto a la categoría de ingresos en la que es clasificada, permite identificar la variación a través del tiempo para los distintos deciles. La Figura 2.4. muestra el porcentaje de hogares con deuda por categoría de ingreso para las distintas versiones de la EFH.



**Figura 2.4:** Hogares con deuda ( % del total de hogares en cada categoría de ingresos.)

La gráfica muestra que el nivel de endeudamiento ha aumentado indistintamente para todos los deciles y que su aumento ha sido sostenido durante el rango de tiempo analizado.

#### 2.3.4. Estudios de análisis empírico

Existe gran y variada literatura, principalmente basada en análisis empíricos con datos de encuestas, que estudia cuales grupos en la sociedad que tienen la mayor probabilidad de

enfrentar dificultades financieras para cumplir con sus compromisos económicos u otros pagos y convertirse en sobre endeudados. Como se mencionó anteriormente y siguiendo el estudio de Anderloni y Vandone [1], el análisis econométrico empírico de encuestas es un grupo de estudios que usan métodos econométricos, tales como regresión, series de tiempo, análisis de corte transversal, entre otros, para probar hipótesis de trabajo sobre un gran número de observaciones de carácter público y de libre acceso.

Kim and Devaney [40] utilizan una base de datos derivada de la encuesta de las finanzas del consumidor patrocinada por la junta de gobernadores de la reserva federal. Utilizando un modelo Probit para estudiar el endeudamiento (factores influyentes) individual a través la variable binaria que representa el saldo pendiente o negativo en la tarjeta de crédito y un modelo de regresión con una variable continua representando el monto de saldo pendiente de la tarjeta de crédito los autores obtienen, para el modelo de regresión lineal un R<sup>2</sup> de 0.3013, lo que implica que solo alrededor de un 30 % de la variación en saldos de tarjeta de crédito pendientes fue explicado por la especificación del modelo.

Del Rio y Young [41] utilizan evidencia de la Encuesta de panel de hogares británica (BHPS) para examinar cómo las actitudes hacia la deuda no asegurada se relacionan con las finanzas del hogar y otras características. Se estima un modelo Logit ordenado para 1995 y 2000 utilizando un indicador auto informado de dificultad financiera como variable dependiente. En cuanto a la bondad de ajuste del modelo estimado se obtuvo un pseudo-R<sup>2</sup> de 0.13 y la cantidad de respuestas que se lograron clasificar correctamente dentro de la muestra fue de 75 %, 44 % y 42 %, siendo la última categoría la de mayor interés puesto que identificaba los hogares con problemas graves.

Worthington [42] examina el rol de la deuda como fuente de estrés financiero en los hogares de Australia. Utilizando 3268 datos de la Encuesta de gastos del hogar formula un modelo Logit para determinar la fuente y la magnitud de aquellos factores que inciden al desarrollo de un estrés financiero. El estudio utiliza la especificación de las percepciones del estrés financiero del hogar como la variable dependiente de una regresión con características demográficas, socioeconómicas y de deuda como variables explicativas. A través de una serie de modelo de probabilidad constante y modelos estimados se obtienen predicciones correctas desde un 59.71 %, hasta un 75.21 % respectivamente, respecto al estrés financiero

como variable dependiente. Los modelos muestran valores para entre 0.052 a 0.117 para  $R^2$ . El autor indica que los porcentajes de predicción correcta corresponden a datos “en la muestra” y que el desempeño de los modelos podría variar para datos fuera de la muestra.

Stone y Maury [43] se enfocan en diseñar un modelo de sobre-endeudamiento para ser utilizado en la identificación temprana de miembros de la Fuerza Aérea de US que presenten el mayor riesgo de desarrollar problema en el manejo de las finanzas personales. Utilizando los datos de una encuesta diseñada para que recoge información demográfica, económica, sociológica, psicológica y situacional de 501 individuos, los autores estiman un modelo Logit de endeudamiento del consumidor utilizando deuda personal no garantizada versus ninguna deuda. El modelo exhibió un pseudo- $R^2$  de 0.3329, una sensibilidad y especificidad de 60.84 % y 90.17 % respectivamente y una asertividad total de 80.59 % en la muestra.

En lo que respecta a América Latina, y solo hasta unos años atrás, la información financiera detallada necesaria para realizar investigaciones relevantes en materia de economía de hogares era escasa. Una serie de encuestas de índole financiera han sido desarrolladas y aplicadas en la región gracias al liderazgo del Banco de desarrollo de América Latina (CAF), que en algunos países ha contado con el apoyo de bancos centrales y superintendencias. Basado en estas encuestas, diversos estudios han sido desarrollados con el objetivo de enriquecer la discusión y dar lugar a un mayor entendimiento en temas de endeudamiento, claves para promover el desarrollo sostenible en la región [35].

Las investigaciones aplicadas por los bancos centrales de Bolivia, Brasil, Colombia, Ecuador, México, República Dominicana, Perú, Uruguay y el Centro de Estudios Monetario Latinoamericanos (CEMLA) analizan los principales factores determinantes de las decisiones financieras de los hogares relacionadas con la tenencia de productos de ahorro y crédito, y en especial en la diferencia de uso de productos financieros formales o informales. En Chile, las investigaciones también permanecieron limitadas y fue gracias a la nueva Encuesta Financiera de Hogares (EFH), realizada por primera vez en el 2007 por el Banco Central de Chile que se contribuyó significativamente con nuevos datos e información financiera a través de múltiples variables. Una serie de estudios comenzaron a presentarse abarcando variadas aplicaciones de investigación como estudios de distribución de activos

y deuda de los hogares, mediciones de riesgo financiero de hogares [44] efectos de los ciclos de desempleo en el riesgo de la deuda [45], restricciones de endeudamiento [46] e incumplimiento crediticio [47].

Ruiz-Tagle y Vella [46], con el objetivo de examinar los determinantes de la deuda de los hogares chilenos. exploran las relaciones entre la probabilidad y el nivel de deuda según las características del hogar estimando una serie de modelos destacando un modelo que explica la probabilidad de que un hogar individual esté endeudado y un modelo que explica probabilidad de demanda de crédito. Ambas estimaciones bajo el principio de mínimos cuadrados. El primer modelo presenta una capacidad explicativa del modelo es muy reducida ( $R^2$  de 0.077) y el segundo alcanza un mejor desempeño alcanzado un  $R^2$  de 0.251.

Ruiz-Tagle, Garcia y Miranda [48] analizan la evolución y los determinantes de los indicadores de endeudamiento y sobre endeudamiento, utilizando la EFH y EPS . El estudio también explora el rol de shocks (i.e salud. educación e ingreso) y de las preferencias de los individuos en el sobre endeudamiento de los hogares. Los autores estiman en primera instancia una serie de modelos lineales que buscan explicar el nivel de endeudamiento y posteriormente, una serie de modelos binarios para identificar y estudiar los determinantes del sobre-endeudamiento. En el caso de las estimaciones lineales, los autores destacan la baja capacidad explicativa ( $R^2$  entre 0.001 y 0.05). Para las estimaciones probabilísticas binarias la situación no es muy distinta y los modelos no logran una capacidad explicativa superior (pseudo- $R^2$  por debajo del 0.0857).

Madeira (2013) utiliza las encuestas Casen y EFH para estudiar cómo distintos factores de riesgo afectan la morosidad de las familias a través del ciclo económico entre 2006 y 2011. El método econométrico empleado por el autor consiste en una estimación de dos etapas, donde inicialmente se estima el acceso al crédito y la selección de un determinado monto de deuda para luego estimar la decisión discreta de incumplir o no el préstamo adquirido. Para la primera etapa, la estimación Probit utilizada para estimar la decisión de adquirir o no un crédito presenta una baja capacidad explicativa para los distintos años analizados (un mínimo y un máximos de pseudo- $R^2$  de 0.049 para 2008 y 0.081 para 2007 respectivamente). Respecto a la segunda etapa, la estimación Probit empleada para la

decisión de incumplimiento en el pago de la deuda adquirida, se encuentra una capacidad explicativa de los modelos, un poco superior a la primera etapa. En efecto, los pseudo-R<sup>2</sup> siguen siendo muy bajos pero superan el 10 % en el caso del año 2011 (con un mínimo de 0.055 en 2007).

Muchas veces las estimaciones realizadas con las técnicas econométricas clásicas presentan capacidades explicativas muy bajas, ya sea considerando R<sup>2</sup> o pseudo-R<sup>2</sup> según corresponda. La baja capacidad explicativa de un modelo tiene directa implicancia en la capacidad predictiva del mismo, puesto que esta será muy limitada cuando se busque predecir correctamente nuevos datos. Esta deficiencia en la capacidad predictiva es la motivación principal de esta investigación. Específicamente, se busca aplicar nuevas técnicas de clasificación, en especial técnicas de Inteligencia Artificial, que ayuden a mejorar la eficiencia en la predicción de nuevos datos.

## 2.4. Algoritmos de aprendizaje inteligente

En la actualidad, la presencia de herramientas computacionales en el estudio del sector económico y financiero es amplia y variada. Bahrammirzaee [6] en su revisión reúne las aplicaciones de la inteligencia artificial a temas como, calificación de créditos, optimización de portafolio, estrés financiero y bancarrota.

De acuerdo con nuestra revisión de la literatura, el sobreendeudamiento de hogares es un tema que no ha sido estudiado con herramientas derivadas de la inteligencia artificial. Sin embargo múltiples técnicas computacionales han sido incluidas en el análisis de problemas que consideramos “matemáticamente” similares. Como lo expresan Peng, Wang, Kou y Shi [49] en su investigación, gran parte de los estudios asociados a la predicción de comportamientos financieros con algún grado de riesgo son principalmente “problemas de clasificación binaria” (e.i tras su análisis, las observaciones se clasifican en uno u otro estado). Este artículo considera el sobreendeudamiento y su predicción como un problema del tipo binario donde se presentan dos estados posibles (i.e estar o no estar sobre endeudado). Bajo este punto de vista consideramos principalmente los problemas de predicción de la bancarrota como guía y motivación en cuando al uso de técnicas

inteligentes en la aplicación de problemas financieros binarios.

Trabajar en la predicción de la bancarrota a través de modelos de aprendizaje automático (e.i técnicas de inteligencia artificial) ha permitido mejorar la precisión de predicción, en comparación con métodos econométricos, principalmente gracias a que logran interpretar eficazmente muchas de las relaciones no lineales existentes. Redes neuronales artificiales ([50]; [51]), support vector machine [52], son dos de las técnicas computacionales más utilizadas para el diseño de modelos de bancarrota. Sin embargo, durante los últimos años, la aplicación individual de cada técnica ha quedado atrás y la tendencia en la investigación se ha inclinado a la combinación de dos o más de estas técnicas con el objetivo de mejorar cada vez la precisión alcanzada en la predicción de cada modelo propuesto.

Dentro de las publicaciones más relevantes para el diseño de nuestro trabajo mencionamos a Chandra, Ravi y Bose [53], quienes desarrollan un modelo híbrido basado en las técnicas perceptrón multicapa (MLP), bosque aleatorio (RF), regresión logística (LR), support vector machine (SVM), árboles de clasificación y regresión (CART), para la predicción de la bancarrota en compañías puntocom. Los autores emplean las técnicas de ensamble y boosting para mejorar la asertividad en la predicción. Aplicando PCA para la selección de variables, encuentran los mejores resultados en la combinación de las técnicas RF, SVM y MLP. Comparan los resultados de la técnica propuesta con RF, LR, MLP, SVM, CART concluyendo que su método es superior a todas las técnicas.

Fedorova, Gilenko y Dovzhenko [54] proponen un sistema híbrido mediante las técnicas de análisis discriminante multivariante (MDA), regresión logística (LR), árbol de clasificación y regresión (CRT) y dos conocidos tipos de red neuronal artificial; perceptrón multicapa (MLP) radial basis function network (RBFN), para el análisis de la predicción de bancarrota en empresas rusas. Los mejores resultados fueron obtenidos al combinar las técnicas MLP, CRT y LR.

Azayite y Achchab [55] combinan redes neuronales (ANN) y análisis discriminante (DA) para el análisis de riesgo de quiebra en una muestra de empresas marroquíes. Los autores encuentran que su modelo alcanza una asertividad más alta en comparación a la técnica DA.

Mselmi, Lahiani y Hamza [56] proponen un modelo híbrido basado en las técnicas de

support vector machine (SVM) y los mínimos cuadrados parciales (PLS) para predecir el estrés financiero de empresas francesas. El modelo propuesto es comparado con LR, PLS y ANN superando en asertividad a todas las técnicas.

Siguiendo estas investigaciones, este trabajo utiliza la innovadora combinación entre las técnicas ANFIS y modelo Probit, aplicado a una muestra de hogares chilenos, con el objetivo de predecir su condición de sobreendeudamiento. Consideramos importante destacar que, a nuestro conocimiento, esta combinación no ha sido probada con anterioridad. Comparamos el desempeño del modelo híbrido propuesto con las técnicas Probit, MLP, SVM y ANFIS y los resultados obtenidos cumplen las expectativas de asertividad y nos permiten proponer una mejorada metodología para la identificación preventiva de hogares en riesgo financiero.

### 2.4.1. Modelos de datos desbalanceados

Similar a problemas financieros tales como el fraude y bancarrota, el sobreendeudamiento es un fenómeno poco habitual. Aunque el número de hogares que actualmente vive problemas o riesgo financiero es alta, proporcionalmente los casos de sobreendeudamiento son bajos. De acuerdo con Veganzones y Séverin [57], gran número de modelos de predicción, en especial de predicción de bancarrota, son diseñados bajo “el paradigma clásico de muestras pareadas de datos disponibles”. Considerar un conjunto de datos balanceado (igual cantidad de observaciones para las categorías) puede no ajustarse a la realidad del fenómeno en cuestión, y en consecuencia afectar el modelo y sus resultados.

Un conjunto de datos que presenta una distribución desigual entre sus clases es lo que se conoce como datos desbalanceados, donde se distingue claramente una clase mayoritaria (aquella con mayor cantidad de observaciones) y una clase minoritaria (aquella con menor cantidad de observaciones). Para algunos algoritmos inteligentes, manejar adecuadamente el problema de balanceo de datos puede resultar en un verdadero desafío.

A la fecha se han desarrollado distintos métodos que ayudan a solucionar parcial o totalmente el problema de datos desbalanceados. Su objetivo es evitar que el dominio de una clase sobre otra y de esta forma afectar el desempeño en la predicción. Los métodos encontrados en la literatura se distinguen principalmente por la forma en que manejan la

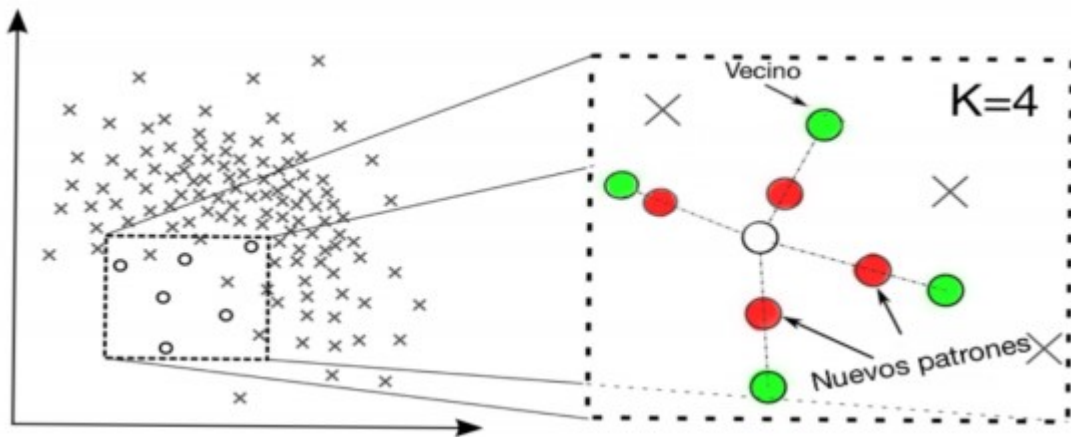
desigualdad numérica entre las clases, esto es; por sobre muestreo (agregar datos) de la clase minoritaria o sub muestreo de la clase mayoritaria (reducción o eliminación de datos).

Para el desarrollo de esta investigación nos enfocamos en los métodos de sobre muestreo, en particular, la técnica de sobre muestreo de minorías sintéticas (Smote), desarrollada por Chawla, Bowyer, Hall, y Kegelmeyer [58]. A través de esta técnica se busca generar un mayor número de observaciones de la clase minoritaria (hogares con sobreendeudamiento) mediante la creación de observaciones “sintéticas” que son el resultado de operaciones sobre los datos reales. Con este método buscamos facilitar un mejor funcionamiento del algoritmo inteligente ANFIS utilizado en el modelo híbrido.

#### 2.4.1.1. Syntetic Minority Over-sampling Technique

Esta técnica consiste en generar “sintéticamente” nuevas observaciones de la clase minoritaria interpolando los valores de las instancias minoritarias más cercanas.

El algoritmo Smote comienza con el principio del clasificador KNN encontrando los K vecinos más cercanos con base en alguno de ellos, de esta forma crea un nuevo patrón sintético (extrapolando). La Figura.2.5 explica el funcionamiento.

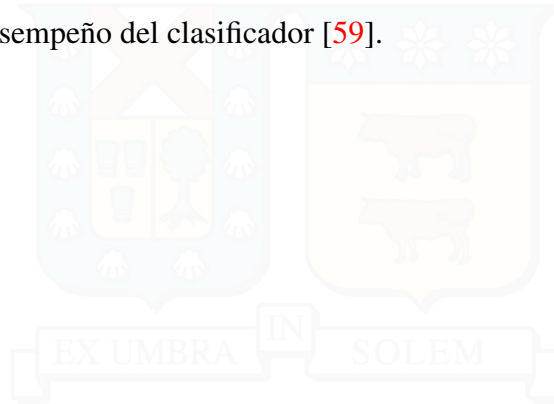


**Figura 2.5:** Funcionamiento SMOTE

El caso de ejemplo mostrado en la imagen anterior refleja la búsqueda que realiza el algoritmo de los K vecinos más cercanos, en este caso los cuatro vecino más cercanos (K=4). Una vez seleccionado los K vecinos más cercanos, el algoritmo selecciona de forma aleatoria uno de los vecinos y aplica la extrapolacion para generar un nuevo patrón en

medio del vecino y el patrón principal. Este procedimiento se repite las veces necesarias hasta que se obtiene un balance deseado.

La principal ventaja de utilizar la técnica de sobre muestreo Smote es que, al incluir observaciones sintéticas se fortalece la región de decisión de la clase minoritaria influyendo positivamente en el desempeño del clasificador [59].



## CAPÍTULO 3

## METODOLOGÍA

### 3.1. Modelos propuestos

#### 3.1.1. Modelo Probit

En estadística, los modelos de elección binaria representan aquellas situaciones en la que individuos son categorizados entre dos alternativas en base a características observables. El rasgo principal de estos modelos es que la variable dependiente es binaria; es decir, solo toma valores entre cero y uno.

La literatura presenta un planteamiento general para los modelos de elección binaria de la forma:

$$Z_i = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + v \quad (3.1)$$

$$Y_i = \begin{cases} 1, Z_i > 0 \\ 0, Z_i < 0 \end{cases} ,$$

$$\begin{aligned}
 Prob(Y_i = 1) &= P \geq 0 \\
 &= P(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + v_i \geq 0) \\
 &= P(v_i \geq 0 - \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k) \\
 &= 1 - P(v_i \leq -\beta_0 - \beta_1 X_1 - \dots - \beta_k X_k)
 \end{aligned} \tag{3.2}$$

Donde:

- $Z_i$ = Variable latente no observable
- $Y_i$ = Variable dicotómica
- $\beta_0$ = Término constante
- $\beta_1 \dots \beta_k$ = Coeficientes del modelo
- $X_1 \dots X_k$ = Variables independientes
- $v_i$ = Término de error

Por otro lado, ya que los modelos de elección binaria estiman probabilidades entre cero y uno, se utiliza una función de densidad acumulada  $F(\cdot)$ , para que los valores predichos se encuentren en ese rango.

$$P(Y = 1) = 1 - F(v_i \leq -\beta_0 - \beta_1 X_1 - \dots - \beta_k X_k) \tag{3.3}$$

$$E(Y) = P(Y = 1) = F(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k) \tag{3.4}$$

Por lo tanto, las características y especificaciones del modelo dependerán del supuesto que se haga respecto a la Función de Distribución Acumulada (FDA) que acumula el término de error ( $v_i$ ). En particular, si se supone distribución uniforme, entonces el modelo será lineal en su probabilidad, si se asume normal será un modelo Probit, mientras que si se elige logística, será un modelo Logit.

El modelo Probit [60] es un tipo de modelo de elección binaria. El objetivo de este modelo es estimar la probabilidad de que una observación con características particulares entre en una u otra categoría previamente definidas.

A continuación se detalla la especificación y características generales del modelo Probit, partiendo desde el esquema general de un modelo de elección binaria.

La función de densidad de probabilidad de la distribución normal es:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (3.5)$$

$$-\infty < x < \infty,$$

Donde:

- $\mu$ =Esperanza
- $\rho$ =Desviación estándar

La función de densidad acumulada es:

$$F(x) = \varphi(x) = \int_{-\infty}^x \varphi(t)dt \quad (3.6)$$

En la que:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sim N(0, 1) \quad (3.7)$$

De esta manera, reemplazando la Ecuación 3.7 en la Ecuación 3.3 se obtiene la expresión final del modelo Probit

$$Y = F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = F(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k) \quad (3.8)$$

Para la predicción del modelo nos basamos en los coeficientes estimados en la muestra, utilizando como punto de corte el valor superior o igual a 0.5 es decir:

Si  $y^* \geq 0.5$ , entonces el hogar está sobre endeudado

Si  $y^* < 0.5$ , entonces el hogar no está sobre endeudado

donde la variable  $y^*$  indica los valores predichos. El principal problema con esta medida es que la elección del punto de corte es arbitraria. Tradicionalmente, se ha utilizado un punto de corte 0.5. Sin embargo, no hay ninguna razón por la cual este corte sea el adecuado.

### 3.1.2. Red neuronal

Una red neuronal artificial (ANN) es un mecanismo de procesamiento de información desarrollado para imitar los sistemas neuronales biológicos básicos, el cerebro humano en particular.

#### 3.1.2.1. Estructura de una Red Neuronal Artificial

La estructura básica de una Red Neuronal se compone de:

- **Neurona:** La unidad fundamental de una RNA se llama neurona artificial y es un emulador básico de la neurona biológica. Esta neurona artificial recibe un conjunto de entradas de información, cada entrada de información es ponderada según el grado de conexión de la sinapsis (también conocido como el peso de cada entrada) y posteriormente se suman, determinando así el nivel de excitación o activación de la neurona.
- **Capas:** Como se ha mencionado anteriormente, una RNA esta conformada por un gran número de neuronas artificiales y por ende, resulta esencial agruparlas según sea su función o tipo. La literatura agrupa a las neuronas en 3 conjuntos.

Capa de Entrada (input layer): Conjunto de neuronas que reciben datos del entorno.

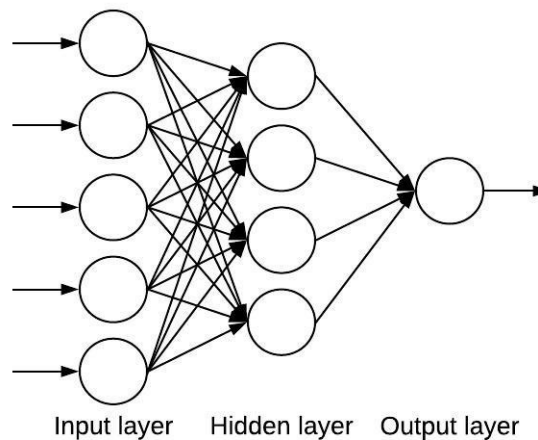
Capa Oculta (hidden layer): Conjunto de neuronas que no tienen conexión directa con el entorno y además procesan toda la información, usualmente una red posee más de una capa oculta.

Capa de Salida (output layer): Conjunto de neuronas que fabrican la salida efectiva de la red neuronal.

En esta investigación nos centraremos en una estructura particular de ANNs, el perceptrón multicapa (MLP) ([61];[62]).

### 3.1.3. Multilayer Perceptron

Un MLP normalmente está compuesto por varias capas de nodos. La primera es una capa de entrada donde se recibe información externa. La última capa es una capa de salida donde se obtiene la solución problema. La capa de entrada y la capa de salida están separadas por una o más capas intermedias llamadas capas ocultas. Los nodos en las capas adyacentes están generalmente conectados por arcos acíclicos desde una capa previa a una capa siguiente. La Figura.3.1. muestra un ejemplo de MLP totalmente conectado con una capa oculta.



**Figura 3.1:** Arquitectura Perceptrón Multicapa

La elaboración de una red neuronal para un problema de particular es una tarea no trivial. Existen variables en el modelamiento de una red que afectan el rendimiento de la misma, por lo que deben considerarse cuidadosamente. Una decisión crítica en la determinación de la arquitectura apropiada es número de capas, y el número de nodos en cada capa. La selección de estos parámetros es básicamente dependiente del problema. Aunque existen muchos enfoques diferentes para determinar la arquitectura óptima de una ANN, estos métodos no pueden garantizar la solución óptima para todos los tipos de problemas [63].

En la red que utilizaremos en esta investigación, la capa de entrada tendrá  $p$  elementos de procesamiento, es decir, uno para cada variable de predicción. Cada elemento de procesamiento en la capa de entrada envía señales  $x_i$  ( $i = 1 \dots p$ ) a cada uno de los elementos de procesamiento. Cada elemento de procesamiento en la capa oculta (indexado por  $j = 1 \dots q$ ) produce una “activación” ( $a_j$ ).

$$a_j = \mathbf{G} \left( \sum_i w_{ij} x_i \right) \quad (3.9)$$

donde  $w_{ij}$  son los pesos asociados con las conexiones entre los elementos de procesamiento  $p$  de la capa de entrada y el elemento de procesamiento  $j$ -ésimo de la capa oculta.

Las conexiones entre nodos en una red determinan fundamentalmente el comportamiento esta. La función de activación o de transferencia determina la relación entre las entradas y salidas de un nodo y una red. En la práctica, solo se usa un pequeño número de funciones de activación como la tangente sigmoidea, la tangente hiperbólica y las funciones sinusoidales, esto debido a sus características no lineales y diferenciables.

$$\mathbf{G}(x) = \begin{cases} 1/(1 + e^{-x}) \\ (e^x - e^{-x})/(e^x + e^{-x}) \\ \sin x \end{cases} \quad (3.10)$$

No está claro si las diferentes funciones de activación tienen efectos importantes en el rendimiento de las redes, sin embargo, Klimasauskas [64] sugiere la función de activación logística para los problemas de clasificación, donde los valores objetivos son usualmente binarios.

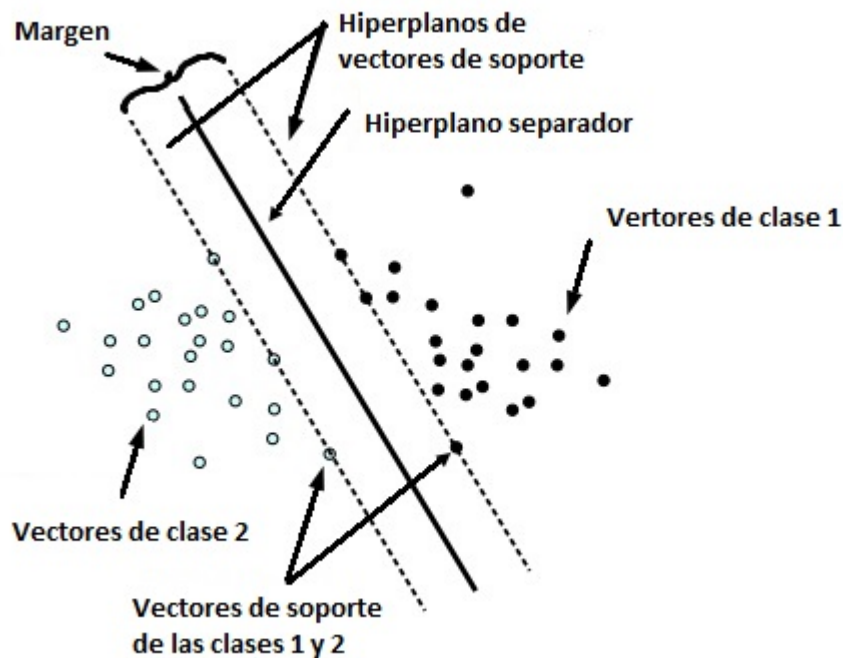
Los pesos de arco de una red se modifican iterativamente con el objetivo es encontrar valores adecuados para minimizar la desviación entre la salida dada por la red y la salida deseada. Este proceso se conoce como entrenamiento, siendo el método más conocido el de retro-propagación [65].

El método de retro-propagación utiliza el descenso de gradiente para cambiar los pesos proporcionales a la derivada de la función de error con respecto a cada peso. Para

el algoritmo de descenso de gradiente, se deben especificar los parámetros de tasa de aprendizaje y momentum, que influyen en la magnitud de los cambios de peso.

### 3.1.4. Support Vector Machine

Support Vector Machine (SVM) desarrollada Vapnik [66] surge como una poderosa herramienta de clasificación que se basa en el principio de margen máximo a partir de los vectores de soporte. El hiperplano de separación entre clases se obtiene a partir de la solución de un problema de optimización el cual se interpreta como la distancia máxima entre los hiperplanos que contienen los vectores de soporte de ambas clases (márgen máximo). La Figura 3.2 representa el funcionamiento de SVM a través de sus conceptos básicos.



**Figura 3.2:** Support Vector Machine

Matemáticamente, SVM puede ser descrito de la siguiente manera: para un conjunto de entrenamiento dado, denotado por  $S = \{(\mathbf{x}_i, \mathbf{t}_i) \mid i = 1, 2, \dots, n\}$ , donde  $\mathbf{x}_i \in R^m$  y  $\mathbf{t}_i \in \{-1, 1\}$ , después de mapear el conjunto de entrada  $\mathbf{x}$  en un espacio de característica de alta dimensión, los datos de entrenamiento pueden separarse mediante un conjunto de hiperplanos lineales.

El problema de optimización que entrega el hiperplano de separación se describe como:

$$\text{mín}(\frac{1}{2}\alpha^T Q\alpha - \sum_{i=1}^n \alpha_i) \quad (3.11)$$

sujeto a

$$0 \leq \alpha_i \leq C$$

$$\sum_{i=1}^n t_i \alpha_i = 0$$

donde  $\alpha_i$  es una multiplicación de Lagrange para la muestra de entrenamiento  $i$ ,  $C$  ( $C > 0$ ) es el límite superior de  $\alpha_i$ ,  $Q$  es una matriz de  $n$  por  $n$  con  $Q_{ij} = t_i t_j k(x_i, x_j)$ , y  $k(x_i, x_j)$  es la función Kernel [66].

Existen una serie de funciones kernel, siendo las más conocidas las lineales, polinomiales, radiales (RBF) y sigmoideas.

- Lineal  $K(x_i, x_j) = x_i^T x_j$
- Polinomial  $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$ ,  $\gamma > 0$
- Radiales (RBF)  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ ,  $\gamma > 0$
- Sigmoideas  $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

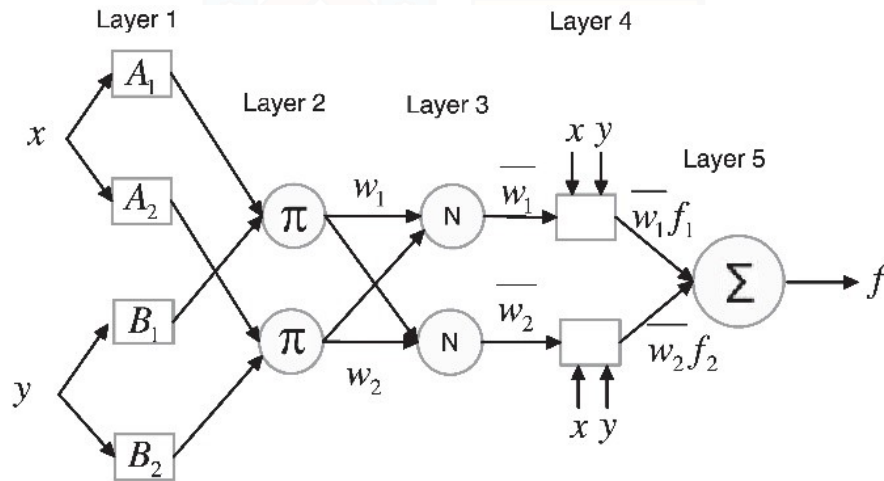
con  $\gamma, r$  y  $d$  son parametros del kernel.

En este trabajo se define el problema de sobreendeudamiento como un problema no lineal y se utiliza la función kernel radial RBF para optimizar el hiperplano. La función kernel radial (RBF) definida anteriormente se utiliza debido a su falta de linealidad, posee un pequeño número de hiperparámetros y menos dificultades numéricas [67] en comparación con otras funciones kernel.

### 3.1.5. Adaptive Neuro Fuzzy Inference System

La técnica Adaptive Neuro-Fuzzy Inference System (ANFIS) fue presentada originalmente por Jang en 1993 [68]. ANFIS es una técnica simple de aprendizaje de datos que utiliza Fuzzy Logic para transformar entradas dadas en un resultado deseado a través de

elementos de procesamiento de red neuronal altamente interconectados y conexiones de información, que se ponderan para asignar las entradas numéricas a un resultado. ANFIS combina los beneficios de dos técnicas de aprendizaje automático (Fuzzy Logic y Neural Network) en una sola técnica la aplicación de métodos de aprendizaje de la red neuronal para ajustar los parámetros de un sistema de inferencia difusa (FIS).



**Figura 3.3:** Arquitectura ANFIS

Para presentar la arquitectura ANFIS, se consideran dos reglas difusas IF-THEN basadas en un modelo Sugeno de primer orden:

$R_1$ : IF  $x$  is  $A_1$  AND  $y$  is  $B_1$  THEN

$$f_1 = p_1x + q_1y + r_1$$

$R_2$ : IF  $x$  is  $A_2$  AND  $y$  is  $B_2$  THEN

$$f_2 = p_2x + q_2y + r_2$$

donde  $x$  e  $y$  son variables independientes,  $A_i$  y  $B_i$  son conjuntos difusos (etiquetas lingüísticas como: pequeño y grande),  $p_i$ ,  $q_i$ ,  $r_i$  son los parámetros de la variable dependiente. La Fig. 3.3. Ilustra el mecanismo de razonamiento para este modelo Sugeno, que es la base del modelo ANFIS. Círculos indican un nodo fijo, mientras que los cuadrados indican un nodo adaptativo. ANFIS tiene una arquitectura de cinco capas. Cada capa se explica en detalle a continuación.

- Capa 1: En esta capa todos los nodos son nodos adaptativos y los resultados de esta capa son el grado de membresía difusa de las entradas, que están dadas por las siguientes ecuaciones:

$$O_{1,i} = \mu_{A_i}(x) ; i=1,2.$$

$$O_{1,i} = \mu_{B_{i-2}}(y) ; i=3,4.$$

donde  $x$  e  $y$  son las entradas del nodo  $i$ , y  $A_i$  y  $B_i$  son las etiquetas lingüísticas asociadas con esta función de nodo,  $\mu_{A_i}(x)$  y  $\mu_{B_{i-2}}(y)$  pueden adoptar cualquier función de membresía fuzzy (por ejemplo triangular, trapezoidal, campana, etc.).

- Capa 2: Los nodos son nodos fijos. Esta capa implica operadores difusos; usa el operador AND para confundir las entradas. Están etiquetados con  $\pi$ , lo que indica que funcionan como un simple multiplicador. La salida de esta capa se puede representar como:

$$O_{2,i} = \omega_i = \mu_{A_i}(x) * \mu_{B_i}(y) ; i=1,2.$$

Estas son las llamadas “fuerzas de disparo” de las reglas.

- Capa 3: Los nodos también son nodos fijos etiquetados por  $N$ , para indicar que ellos juegan un rol de normalización a las denominadas “fuerzas de disparo” de la capa anterior. La salida de esta capa puede ser representada como:

$$O_{3,i} = \bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2} ; i = 1,2.$$

Las salidas de esta capa son llamadas fuerzas de disparo normalizadas.

- Capa 4: Los nodos son adaptativos. La salida de cada nodo es el producto entre la fuerza de disparo normalizada y un polinomio de primer orden (por un modelo Sugeno de primer orden). La salida de la capa está dada por:

$$O_{4,i} = \bar{\omega}_i f_i = \bar{\omega}_i (p_i x + q_i y + r_i) ; i = 1,2.$$

Donde  $\omega_i$  es la salida de las Capa 3, y  $p_i$ ,  $q_i$  y  $r_i$  son los respectivos parámetros.

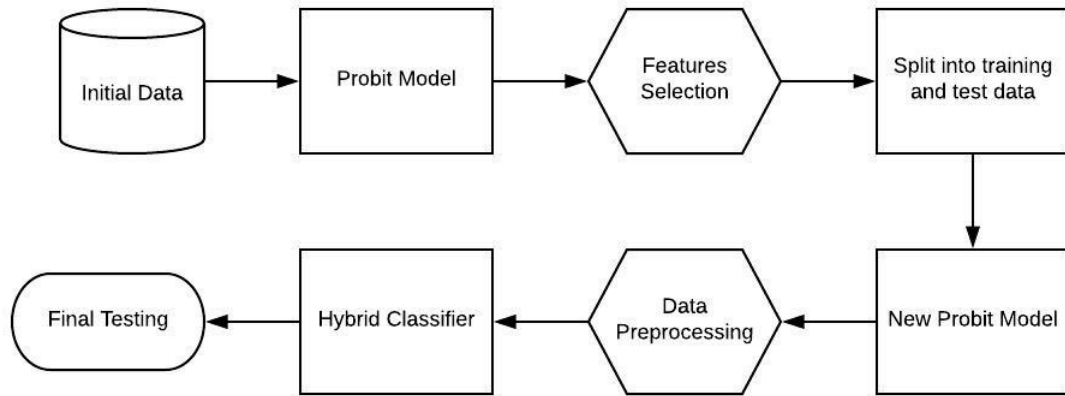
- Capa 5: En esta última capa hay solo un nodo fijo etiquetado con  $\Sigma$ . Este nodo realiza la suma de todas las señales entrantes. La producción general del modelo está dada por:

$$O_{5,i} = \sum_i \bar{\omega} f_i = \frac{\sum_i \omega_i f_i}{\sum_i \omega_i}.$$

### 3.1.6. Modelo Híbrido

Este trabajo propone un modelo híbrido basado en Adaptive Neuro Fuzzy Inference System en conjunto con el modelo Probit. Ambas metodologías se combinan para la clasificación y predicción del sobreendeudamiento en hogares. Para este propósito se utiliza el algoritmo ANFIS como clasificador y un modelo Probit como variable de entrada a este clasificador. El procedimiento del modelo híbrido propuesto se muestra en la Figura.3.4.

Como se observa en la representación, previo a la elaboración del modelo se determinan aquellas variables significativas en base al análisis de significancia arrojado por el modelo base Probit al total de los datos analizados (3425 datos). Solo siete variables resultan significativas; Edad, Trabajadores, Tarjeta, Ingreso, Nivel educacional, Carga hipotecaria, Deuda educacional. Utilizamos estas siete variables como entradas para los modelos individuales Probit, MLP, SVM y ANFIS. Para el modelo híbrido propuesto se agrega una octava entrada correspondiente a los resultados obtenidos del modelo Probit aplicado a los datos de entrenamiento (2740 datos). El algoritmo de nuestro modelo híbrido propuesto es el siguiente:



**Figura 3.4:** Proceso de elaboración del modelo híbrido propuesto

- Step 1: Evaluación de un modelo Probit sobre el total de los datos y variables existentes.
- Step 2: Determinación de las variables significativas según el modelo Probit obtenido en el paso anterior.
- Step 3: Reconstrucción del conjunto de datos y realizar una partición aleatoria en datos de entrenamiento (80 %) y datos de prueba (20 %).
- Step 4: Evaluación de un modelo Probit al conjunto de datos de entrenamiento.
- Step 5: Agregar el resultado del modelo Probit generado en el paso 4 como variable de entrada y aplicar el pre procesamiento de datos al nuevo conjunto de variables (Smote y escalamiento).
- Step 6: Evaluación del algoritmo ANFIS al conjunto de datos obtenidos del paso anterior.
- Step 7: Evaluación del modelo híbrido Probit-ANFIS al conjunto de datos de entrenamiento.

## 3.2. Data

### 3.2.1. Descripción de datos

La Encuesta Financiera de Hogares 2014 (EFH) es un estudio que lleva a cabo el Banco Central de Chile desde el año 2007 y tiene como objetivo generar información detallada del balance financiero de los hogares, la que no está disponible en otras fuentes de datos públicos. Para la realización de este estudio se trabajó sobre los datos obtenidos en la última versión de la encuesta realizada el 2014, considerando 3425 hogares<sup>5</sup>.

La Tabla 3.1 muestra las variables seleccionadas<sup>6</sup> para el diseño del modelo y sus estadísticas resumidas.

Variables		Media	Std. Dev	Min	Max
$X_1$	Género	0.33	0.47	0	1
$X_2$	Edad	50.05	15.62	18	90
$X_3$	Ahorro	0.45	4.35	0	180
$X_4$	Trabajadores	1.49	0.97	0	9
$X_5$	Tarjeta	0.25	0.43	0	1
$X_6$	Estado Civil	0.62	0.48	0	1
$X_7$	Ingreso	1.88	7.67	0.018	371.66
$X_8$	Nivel Educativo	12.89	4.16	0	29
$X_9$	Carga Hipotecaria	0.04	0.18	0	4.48
$X_{10}$	Deuda Educativa	0.07	0.97	0	32.42
Y	Sobreendeudamiento	0.09	0.28	0	1

TABLA 3.1: Estadísticas descriptivas.

Las variables incluidas en este estudio se derivan de la literatura existente en materia

<sup>5</sup>La EFH 2014 alcanzó un total de 4502 observaciones correspondientes a hogares de todo el país. Se eliminaron aquellas observaciones que no presentaban información en las variables seleccionadas para el modelo logrando de esta forma obtener un total de 3425 observaciones sin datos perdidos.

<sup>6</sup>La información contenida en las variables  $X_1, X_2, X_5, X_6$  y  $X_8$  representan datos de aquellos individuos señalados como jefes de hogar. Más detalles de los datos y variables utilizadas en el Apéndice 5.1

de sobre endeudamiento de hogares y representan información referente a cada hogar. Del total de datos, un 67 % de los hogares presentan un jefe de hogar hombre y un 33 % una mujer; con una edad promedio de 50 años. En promedio existen 1.5 trabajadores por cada hogar y que los ingresos alcanzan en promedio los MM\$ 1.88 por hogar (aproximadamente USD\$2.990). Más detalles de cada variable se exponen en el Apéndice 5.1.

De acuerdo a lo descrito en la sección 3.1.6, se realiza una selección de variables según el nivel de significancia obtenido en un modelo Probit inicial sobre la totalidad de observaciones. Tabla 3.1 muestra las variables y su respectivo nivel de significancia. Para la evaluación de todos los modelos considerados en este artículo se utilizan siete variables finales.

### 3.2.2. Pre procesamiento de datos

Para el diseño y evaluación de la capacidad predictiva de los distintos modelos consideramos un conjunto de entrenamiento (80 %) con 2740 observaciones y un conjunto de prueba (20 %) con 685 observaciones del total de observaciones disponibles.

Variables		Coefficiente	p-value
$X_1$	Género	0.1072	0.1684
$X_2$	Edad	-1.7829	0
$X_3$	Ahorro	-7.7060	0.1044
$X_4$	Trabajadores	-1.6141	0
$X_5$	Tarjeta	1.0718	0
$X_6$	Estado Civil	-0.0460	0.5657
$X_7$	Ingreso	-86.1028	0
$X_8$	Nivel Educativo	-1.4422	0
$X_9$	Carga Hipotecaria	14.5476	0
$X_{10}$	Deuda Educativa	2.8922	0.010

TABLA 3.2: Selección de variables.

### 3.2.2.1. Smote

Como primer paso en el pre procesamiento de los datos enfrentamos el problema de balance. Como se muestra en la tabla 1, la variable dependiente tiene una media de 0.09, lo que indica solo un 8.98 % de los hogares en la muestra total está sobre endeudado. Misma situación se mantiene para el conjunto de datos de entrenamiento como se muestra en la tabla 4. Para enfrentar esta situación aplicamos la técnica de sobre muestreo Smote desarrollada por Chawla [58] tanto al conjunto de datos completos como a los datos de entrenamiento.

Con la aplicación de esta técnica logramos aumentar la clase minoritaria de un 8.98 % a un 16.45 % al conjunto de entrenamiento.

Conjunto	Original		SMOTE		Total (%)
	Y=1	Y=0	Y=1	Y=0	
Full data	8.82	91.18	15.24	84.76	100
Training	8.98	91.02	16.45	83.55	80
Test	8.18	91.82	-	-	20

TABLA 3.3: Distribución de la variable independiente (%).

### 3.2.2.2. Escalamiento de datos

Un segundo paso en el pre procesamiento de datos es el escalamiento de los datos de entrada. Escalar datos antes de la aplicación de un algoritmo inteligente es muy importante [69]. El beneficio es evitar los atributos en mayores rangos numéricos que dominan aquellos en rangos numéricos más pequeños. Otra ventaja es que se evitan las dificultades numéricas durante el cálculo de los distintos modelos. Utilizamos el método de escalado lineal para cada atributo en el rango [0.1] según la ecuación 3.12. El mismo método para escalar se usa para los datos de entrenamiento y prueba.

$$\mathbf{X}' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.12)$$

# CAPÍTULO 4

## RESULTADOS

### 4.1. Resultados

En esta sección se detallan y discuten los resultados obtenidos por todos los modelos propuestos. Se compara con los resultados que entrega el modelo híbrido con los modelos Probit, MLP, SVM y ANFIS.

#### 4.1.1. Resultados modelo Probit

Tal como se señala previamente, del conjunto definido como datos de prueba, un total de 56 observaciones representan hogares con sobreendeudamiento (equivalente a un 8.18 %) y 629 observaciones corresponden a hogares sin sobreendeudamiento.

Los resultados de predicción y la matriz de confusión para el modelo Probit sobre los datos de prueba se muestran en la Tabla 4.1.

		Estado predicho		
		Estado de sobreendeudamiento	Sobre endeudado	No Sobre endeudado
Estado actual	Número de muestras	Sobre endeudado	30	26
		No sobre endeudado	28	601

TABLA 4.1: Resultados modelo de sobreendeudamiento con Probit.

Un total de 30 hogares son correctamente clasificados como sobre endeudados lo que corresponde a un 53.57 % y 26 hogares con erróneamente clasificados lo que corresponde a un 46.43 %.

#### 4.1.2. Resultados modelo MLP

El número de nodos en las capas de entrada y salida ha sido establecido de acuerdo a la estructura del análisis, es decir, el número de variables explicativas y de salida, respectivamente.

Por otro lado, se llevaron a cabo varios experimentos para encontrar la cantidad de elementos ocultos que proporcionaban la mayor precisión en la predicción del conjunto de datos de prueba. La Tabla 4.2 muestra los resultados (porcentaje promedio de casos correctamente clasificados) obtenidos a partir de seis configuraciones diferentes, variando la cantidad de elementos ocultos de 2 a 7 y entrenando diez veces cada topología.

El rango de variación para el número de elementos ocultos es definido de acuerdo a la regla de la pirámide geométrica que se basa en la suposición de que el número de neuronas de la capa oculta ha de ser inferior al total de variables de entrada, pero superior al número de variables de salida [70]. La velocidad de aprendizaje, el momento y las épocas de entrenamiento se establecen en 0.1, 0.8 y 5000, respectivamente. La arquitectura de red con la tasa de precisión promedio más alta se considera como la arquitectura de red óptima.

Número de nodos ocultos	ARR
2	77.12
3	76.98
4	77.66(*)
5	75.49
6	76.44
7	76.17

(\*) Mejor arquitectura MLP

TABLA 4.2: Variación en la clasificación y promedio de clasificación según el número de nodos ocultos (%).

Una red 7-4-1 (e.i siete neuronas en la capa de entrada, cuatro neuronas en la capa oculta y una neurona de salida) resulta la arquitectura óptima.

Debido a que hay pocas formas sistemáticas de seleccionar la velocidad de aprendizaje y el momento de forma simultánea, los mejores valores de estos parámetros se eligen a través de la experimentación.

Se realiza una segunda prueba variando los parámetros aprendizaje en la configuración óptima. Si bien una tasa de aprendizaje baja tiende a desacelerar el proceso de aprendizaje, mientras que las tasas de aprendizaje más grandes (especialmente mayores a 0.4), pueden causar oscilación de red y dificultad de convergencia [63], se probaron seis tasas de aprendizaje (de 0.01 a 0.4) y tres valores de momentos (de 0.7 a 0.9). La tasa promedio de precisión de varios parámetros de aprendizaje para la arquitectura 7-4-1 se muestra en la Tabla 4.3.

Tasa de aprendizaje	Momentum		
	0.7	0.8	0.9
0.01	75.55	75.55	78.00
0.05	75.96	78.55	78.70
0.1	78.23	77.66	78.89 (*)
0.2	77.66	77.59	77.73
0.3	77.53	76.51	76.30
0.4	76.23	76.91	76.71

(\*) Mejor arquitectura MLP

TABLA 4.3: Variación en la clasificación promedio según la tasa de aprendizaje y momentum (%)

La matriz de confusión en la Tabla 4.4 muestra los resultados del modelo MLP de arquitectura, con tasa de aprendizaje 0.1 y momentum 0.9, para el conjunto de datos de prueba.

		Estado predicho		
		Estado de	Sobre	No Sobre
		sobreendeudamiento	endeudado	Sobre endeudado
Estado actual	Número de muestras	Sobre endeudado	35	21
		No sobre endeudado	30	599

TABLA 4.4: Resultados modelo de sobreendeudamiento con MLP.

Un total de 35 hogares son correctamente clasificados como sobre endeudados (62.5 %) y 21 hogares son erróneamente clasificados (37.5 %).

### 4.1.3. Resultados modelo SVM

Para mejorar la precisión de clasificación de SVM con kernel RBF, se requiere optimizar el parámetro de costo  $C$  y el parámetro de kernel RBF  $\gamma$  durante la fase de aprendizaje.

Siguiendo las sugerencias de Hsu, Chang and Lin [67] se consideran una validación cruzada de 5 veces (para evitar el problema de sobre clasificación) y se iteraron los parámetros de costo ( $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ ) y  $\gamma$  ( $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$ ).

Después de realizar la grilla de búsqueda de datos de entrenamiento, encontramos los parámetros óptimos  $(C, \gamma) = (2^{14.5}, 2^{0.5})$  y una tasa de validación cruzada de 92.66 %.

		Estado predicho		
		Estado de	Sobre	No Sobre
		sobreendeudamiento	endeudado	Sobre endeudado
Estado actual	Número de muestras	Sobre endeudado	38	18
		No sobre endeudado	32	597

TABLA 4.5: Resultados modelo de sobreendeudamiento con SVM.

La tabla 4.5 muestra la matriz de confusión de los resultados del modelo SVM para el conjunto de datos de prueba. Un total de 38 hogares son correctamente clasificados como sobre endeudados (67.86 %) y 18 hogares con erróneamente clasificados (32.14 %).

#### 4.1.4. Resultados modelo ANFIS

Los datos correspondientes a las siete variables significativas son cargados y se busca determinar los parámetros relevantes para la elaboración del modelo ANFIS.

Respecto a las funciones de membresía, en primera instancia, desarrollamos modelos ANFIS asignando dos funciones de membresía "tri", "trap", "gbell", "gauss", "gauss2z" "pi" para cada entrada, respectivamente.

Posterior a la asignación de funciones de membresía, los modelos ANFIS crean las reglas que se usan para entrenar los modelos. Cada modelo ANFIS fue entrenado con 1000 epoch. No se realizan variaciones adicionales al número de funciones de membresía dado el tiempo de procesamiento excesivo. Los resultados de estos modelos se muestran en la Tabla 4.6.

La arquitectura ANFIS con la tasa de precisión promedio más alta se considera como el modelo óptimo. El modelo ANFIS óptimo con una tasa de precisión promedio del 85.69 % tiene dos funciones de membresía de Gauss en cada entrada.

Tipo de función de membresía	AAR
Tri	85.30
Trap	85.54
Gbell	85.46
Gauss	85.69(*)
Gauss2	85.38
Pi	85.54

(\*) Mejor arquitectura ANFIS

TABLA 4.6: Resultados del modelo ANFIS con distintas funciones de membresía (%).

Como se ve en la Tabla 4.7 que muestra la matriz de confusión de los resultados del modelo ANFIS para el conjunto de datos de prueba; un total de 43 hogares son correctamente clasificados como sobre endeudados (76.79 %) y 13 hogares erróneamente clasificados como no sobre endeudados (23.21 %).

		Estado predicho		
		Estado de	Sobre	No Sobre
		sobreendeudamiento	endeudado	Sobre endeudado
Estado actual	Número de muestras	Sobre endeudado	30	26
		No sobre endeudado	28	601

TABLA 4.7: Resultados modelo de sobreendeudamiento con ANFIS.

#### 4.1.5. Resultados modelo Híbrido

Para el modelo Probit-ANFIS se adiciona una variable extra a las siete variables utilizada en los modelos anteriores que corresponde al resultado obtenido del modelo Probit individual. De forma similar a la etapa anterior, se elaboraron modelos híbridos asignando dos funciones de membresía "tri", "trap", "gbell", "gauss", "gauss2" "pi" para cada entrada, respectivamente. El número de epoch se mantiene en 1000. Resultados de estos modelos se muestran en la Tabla 4.8.

La arquitectura ANFIS con la tasa de precisión promedio más alta se considera como el modelo óptimo. El modelo ANFIS óptimo con una tasa de precisión promedio del 88.37 % tiene dos funciones de membresía de Pi en cada entrada.

Tipo de función de membresía	AAR
Tri	87.71
Trap	87.40
Gbell	88.21
Gauss	87.48
Gauss2	87.71
Pi	88.37(*)

(\*) Mejor arquitectura ANFIS

TABLA 4.8: Resultados del modelo híbrido Probit-ANFIS con distintas funciones de membresía (%).

Como se ve en la Tabla 4.9 que muestra la matriz de confusión de los resultados del modelo ANFIS para el conjunto de datos de prueba; un total de 46 hogares son

correctamente clasificados como sobre endeudados (82.14 %) y 10 hogares erróneamente clasificados como no sobre endeudados (17.86 %).

		Estado predicho		
		Estado de sobreendeudamiento	Sobre endeudado	No Sobre Sobre endeudado
Estado actual	Número de muestras	Sobre endeudado	46	10
		No sobre endeudado	34	595

TABLA 4.9: Resultados modelo de sobreendeudamiento con Probit-ANFIS.

## 4.2. Discusión de resultados

### 4.2.1. Wilcoxon test

La estructura desbalanceada de los datos de sobreendeudamiento obstaculiza la separación de los datos (correcta clasificación) por parte de las distintas técnicas de clasificación debido a la falta de datos para caracterizar adecuadamente la clase de interés. Sin embargo, los resultados obtenidos por los modelos aplicados indican una precisión alta en la clasificación de ambas clases de datos.

Para evaluar el rendimiento de clasificación de cuatro modelos de clasificación de sobreendeudamiento construidos, las tasas de precisión de clasificación y las tasas de precisión promedio se resumen en la Tabla 4.10. Los resultados muestran que el modelo ANFIS híbrido propuesto logra clasificar correctamente un 28.57 %. 14.28 y 5.35 % más de casos de hogares con sobreendeudamiento versus los modelos Probit, MLP, SVM y ANFIS respectivamente.

De forma similar la tasa de precisión promedio es superior a la obtenida por los modelos individuales. Podemos entonces concluir que el modelo de sobreendeudamiento ANFIS híbrido propuesto que se basa en Probit y ANFIS, tiene la mejor capacidad de clasificación de los modelos utilizados como benchmark. Este resultado también se confirmó con una prueba de los rangos con signo de Wilcoxon [71]. Este test no paramétrico permite comparar resultados cuando dos muestras relacionadas no siguen una distribución normal, como es el

caso de los modelos de sobre endeudamiento construidos. Como se puede ver en la Tabla 4.11, los valores p demuestran que el modelo híbrido ANFIS propuesto es claramente mejor que Probit, MLP y SVM, mientras que la diferencia no es significativa cuando se compara con ANFIS.

	$AR_p$	ARR
Probit	53.57	74.56
MLP	62.50	78.89
SVM	67.86	81.38
ANFIS	76.79	85.69
Hibrido ANFIS	82.14	88.37

TABLA 4.10: Tasa de precisión y Tasa de precisión promedio de los modelos construidos (%).

Comparación	p-value	Hipotesis ( $\alpha=0.05$ )
Probit vs MLP	0.173	Acepta
Probit vs SVM	0.224	Acepta
Probit vs ANFIS	0.032	Rechaza
Probit vs Hibrido ANFIS	0.011	Rechaza
MLP vs SVM	0.602	Acepta
MLP vs ANFIS	0.045	Rechaza
MLP vs Hibrido ANFIS	0.024	Rechaza
SVM vs ANFIS	0.047	Rechaza
SVM vs Hibrido ANFIS	0.039	Rechaza
ANFIS vs Hibrido ANFIS	0.209	Acepta

TABLA 4.11: Wilcoxon signed-rank test para comparación de los resultados de clasificación obtenidos.

### 4.2.2. Errores de clasificación

La Tabla 4.12 muestra los errores de Tipo I y Tipo II para los cuatro modelos elaborados. Los resultados muestran que la clasificación errada de hogares con sobreendeudamiento,

representada por el Error Tipo I, registrada por el modelo ANFIS híbrido propuesto es la menor en relación a los modelos Probit, MLP, SVM y ANFIS.

Los niveles del clasificación errada de la clase sobre endeudada logra reducirse considerablemente pero debido a la estructura desbalanceada de los datos, estos niveles siguen siendo altos. Respecto al error Tipo II, el modelo ANFIS híbrido propuesto registra el valor más alto junto al modelo ANFIS individual.

	Error Type I	Error Type II
Probit	46.43	4.45
MLP	37.50	4.77
SVM	32.14	5.09
ANFIS	23.21	5.41
Hibrido ANFIS	17.86	5.41

TABLA 4.12: Error Tipo I y Tipo II de los modelos construidos (%).

## CAPÍTULO 5

## ANEXOS

### 5.1. Descripción de variables

La siguiente tabla describe las variables utilizadas en este estudio para la descripción del sobre endeudamiento financiero de hogares para los modelos elaborados.

Variables	
$X_1$ Género	Variable dicotómica que describe el género del jefe de hogar donde: 1: Mujer / 0: Hombre
$X_2$ Edad	Variable entera que indica la edad del jefe de hogar.
$X_3$ Ahorro	Variable continua que describe el monto declarado total de ahorro que registra el hogar en los últimos 12 meses. Expresada en millones de pesos chilenos.
$X_4$ Trabajadores	Variable entera que describe el número de miembros del hogar que se encuentran trabajando.
$X_5$ Tarjeta	Variable dicotómica que indica si algún miembros del hogar posee tarjeta de crédito con: 1: Si posee tarjeta / 0: No posee tarjeta

---

Variables	
$X_6$ Estado Civil	Variable dicotómica que representa el estado civil del jefe de hogar con: 1: Casado o conviviente / 0: Otro
$X_7$ Ingreso	Variable continua que indica el monto total de ingresos que recibe el hogar mensualmente. Expresada en millones de pesos chilenos.
$X_8$ Nivel Educativo	Variable entera que representa los años de escolaridad del jefe de hogar. El sistema educacional de Chile tiene un promedio de 8 años para educación primaria, 4 años para educación secundaria y 6 para educación superior. La encuesta considera también los estudios de postgrado o segundas profesiones.
$X_9$ Carga Hipotecaria	Variable continua que indica el monto total que el hogar paga mensualmente por deuda hipotecaria (dividendo). Expresada en millones de pesos chilenos.
$X_{10}$ Deuda Educativa	Variable continua que indica el monto total adeudado por el hogar por concepto de educación de sus miembros. Expresada en millones de pesos chilenos.

---

## 5.2. Indicadores de endeudamiento

**Deuda** Por deuda se entenderá a todo préstamo de dinero que la institución bancaria o no bancaria otorgue a la persona natural requirente, tanto para efectos de consumo como hipotecario. Para estos efectos, se considera deuda efectiva la compuesta a partir de saldos del stock de capital más intereses devengados de las obligaciones mantenidas en cada periodo de análisis.

**Carga Financiera** La carga financiera mide el porcentaje del ingreso que es destinado al pago de las obligaciones financieras. Se calcula la carga financiera del individuo  $i$  en el mes  $t$ , utilizando la siguiente fórmula:

$$\text{Carga Financiera}_{it} = \frac{\text{Obligaciones Financieras}_{it}}{\text{Ingreso Mensual}_{it}}$$

Donde las obligaciones son las cuotas mensuales de cada uno de los productos que mantiene el deudor  $i$  al mes  $t$ . Para los créditos de consumo en cuotas e hipotecarios para la vivienda, se cuenta con información del plazo, monto y tasa, a partir de lo cual se calcula la cuota utilizando amortización francesa. Por otro lado, para los créditos rotativos, como las tarjetas de crédito y líneas de crédito que no tienen un plazo definido, se estima la obligación mensual considerando que el deudor cancela su deuda utilizando un crédito de consumo en cuotas con características promedio de mercado para tasa de interés y plazo. Por último, para el cálculo de la carga financiera se considera el ingreso mensual neto de descuentos legales.

**Apalancamiento** El apalancamiento mide el número de ingresos mensuales que un deudor tendría que destinar para saldar sus obligaciones financieras por completo y se calcula mediante la fórmula:

$$\text{Apalancamiento}_{it} = \frac{\sum \text{Obligaciones Financieras}_{it}}{\text{Ingreso Mensual}_{it}}$$

Donde la sumatoria de las obligaciones financieras considera todos los productos que mantenga el deudor  $i$  en el mes  $t$ .

**Razón Carga financiera sobre Ingreso** La Razón Carga financiera sobre Ingreso (RCI) corresponde al cociente entre la carga financiera de la deuda total (incluyendo deuda hipotecaria y deuda no hipotecaria) y el ingreso total efectivo mensual del hogar. Por ingreso total efectivo se entiende el ingreso total mensual que recibe un hogar sin considerar el ingreso imputado a la vivienda principal (valor que los individuos no arrendatarios le asignan a su vivienda por concepto de arriendo).

**Razón Deuda sobre Ingreso RDI** La Razón Deuda sobre Ingreso (RDI) corresponde al cociente entre el valor presente de la deuda total (incluyendo deuda hipotecaria y deuda no hipotecaria) y el ingreso total efectivo anual del hogar.

## Bibliografía

- [1] Luisa Anderloni and Daniela Vandone. Households over-indebtedness in the economic literature. *Universit'a degli Studi di Milano Working Paper*, 46:775, 2008.
- [2] Lungile Ntsalaze and Sylvanus Ikhide. Household Over-indebtedness: Understanding its Extent and Characteristics of those Affected. *Journal of Social Sciences*, 48(1-2):79–93, 2016.
- [3] Gianni Betti, Neil Dourmashkin, Mariacristina Rossi, and Ya Ping Yin. Consumer over indebtedness in the eu: measurement and characteristics. *Journal of Economic Studies*, 34(2):136–156, may 2007.
- [4] Giovanni D'Alessio and Stefano Iezzi. Household over-indebtedness: definition and measurement with italian data. 2013.
- [5] Michal Tkáč and Robert Verner. Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, 38:788–804, jan 2016.
- [6] Arash Bahrammirzaee. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, 19(8):1165–1195, 2010.
- [7] Banco Central de Chile. Informe de estabilidad financiera. Segundo semestre 2017.
- [8] Banco Central de Chile. Cuentas nacionales por sector institucional. evolución del ahorro, la inversión y el financiamiento sectorial en el año 2017. 2017.
- [9] Jonathan Morduch. Income smoothing and consumption smoothing. *Journal of economic perspectives*, 9(3):103–114, 1995.
- [10] James D Gwartney, Randall G Holcombe, and Robert A Lawson. Institutions and the impact of investment on growth. *Kyklos*, 59(2):255–273, 2006.
- [11] Michael Chibba. Financial inclusion, poverty reduction and the millennium development goals. *The European Journal of Development Research*, 21(2):213–230, 2009.
- [12] Rosa Matilde Guerrero Guerrero, Silvia Espinosa, and Kurt Focke. Aproximación a una visión integral de la inclusión financiera: Aspectos conceptuales y casos prácticos. 2010.

- [13] Rahul Anand and Eswar Prasad. Core vs. headline inflation targeting in models with incomplete markets. *Manuscript, Cornell University, Revised version of NBER Working Paper*, (16290), 2012.
- [14] Asli Demirguc-Kunt, Leora Klapper, Dorothe Singer, Saniya Ansar, and Jake Hess. *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. The World Bank, 2018.
- [15] Juan Pablo Montero and Jorge Tarziján. El éxito de las casas comerciales en Chile: ¿regulación o buena gestión? *Documentos de Trabajo (Banco Central de Chile)*, (565):1, 2010.
- [16] Pascaline Dupas, Dean Karlan, Jonathan Robinson, and Diego Ubfal. Banking the unbanked? evidence from three countries. *American Economic Journal: Applied Economics*, 10(2):257–97, 2018.
- [17] Albert Ando and Franco Modigliani. The "life cycle" hypothesis of saving: Aggregate implications and tests. *The American economic review*, 53(1):55–84, 1963.
- [18] Aldo Barba and Massimo Pivetti. Rising household debt: Its causes and macro-economic implications—a long-period analysis. *Cambridge Journal of Economics*, 33(1):113–137, 2008.
- [19] Nathalie Girouard, Mike Kennedy, and Christophe Andre. Has the rise in debt made households more vulnerable? 2006.
- [20] Jessica Schicks. Microfinance Over-Indebtedness: Understanding its drivers and challenging the common myths. *Centre Emile Bernheim (CEB) Working Paper*, 10:47, 2010.
- [21] Merike Kukk. How did household indebtedness hamper consumption during the recession? Evidence from micro data. *Journal of Comparative Economics*, 44(3):764–786, 2016.
- [22] Daniela Fabbri and Mario Padula. Does poor legal enforcement make households credit-constrained? *Journal of Banking & Finance*, 28(10):2369–2397, 2004.
- [23] Stephen E.G. Lea, Paul Webley, and Catherine M. Walker. Psychological factors in consumer debt: Money management, economic socialization, and credit use. *Journal of Economic Psychology*, 16(4):681–701, dec 1995.
- [24] Daniel A Hojman, Álvaro Miranda, and Jaime Ruiz-Tagle. Debt trajectories and mental health. *Social Science & Medicine*, 167:54–62, 2016.
- [25] Steven M Finlay. Predictive models of expenditure and over-indebtedness for assessing the affordability of new consumer credit applications. *Journal of the Operational Research Society*, 57(6):655–669, 2006.
- [26] Katherine Porter. The damage of debt. *Washington & Lee Law Review*, 69:979, 2012.

- [27] John Gathergood. Self-control, financial literacy and consumer over-indebtedness. *Journal of economic psychology*, 33(3):590–602, 2012.
- [28] Mark Bryan, Mark Taylor, and Michail Veliziotis. Over-indebtedness in Great Britain: An analysis using the wealth and assets survey and household annual debtors survey. *Department of Business, Innovation & Skills*, 2010.
- [29] Cristina Ottaviani and Daniela Vandone. Impulsivity and household indebtedness: Evidence from real life. *Journal of Economic Psychology*, 32(5):754–761, oct 2011.
- [30] Jessica Schicks. Over-Indebtedness in Microfinance – An Empirical Analysis of Related Factors on the Borrower Level. *World Development*, 54:301–324, feb 2014.
- [31] Luisa Anderloni, Emanuele Bacchiocchi, and Daniela Vandone. Household financial vulnerability: An empirical analysis. *Research in Economics*, 66(3):284–296, sep 2012.
- [32] Sarah Brown and Daniel Gray. Household finances and well-being in Australia: An empirical analysis of comparison effects. *Journal of Economic Psychology*, 53:17–36, 2016.
- [33] Trust FinMark and Unisa. A Consumer Financial Vulnerability Index for South Africa. 2009.
- [34] Isabelle Guérin, Marc Roesch, Govindan Venkatasubramanian, and Santosh Kumar. Significados múltiples y contradictorios del sobreendeudamiento: Un estudio de caso de hogares en pobreza rural en Tamil Nadu, sur de la India. *Desacatos*, (44):35–50, 2014.
- [35] María José Roa and Diana Mejía. Decisiones financieras de los hogares e inclusión financiera: evidencia para América Latina y el Caribe. 2018.
- [36] William Dudley. Opening remarks at the Convening on Student Loan Data Conference. Technical report, Federal Reserve Bank of New York, 2015.
- [37] Banco de Inglaterra. One Bank Research Agenda. Technical report, Discussion Paper, febrero, 2015.
- [38] Superintendencia de Bancos e Instituciones Financieras Chile. Informe de endeudamiento. Junio 2017.
- [39] Banco Central de Chile. Informe de estabilidad financiera. Primer semestre 2018.
- [40] Haejeong Kim and Sharon A DeVaney. The determinants of outstanding balances among credit card revolvers. *Financial Counseling and Planning*, 12(1):67–77, 2001.
- [41] Ana Del-Río and Garry Young. The impact of unsecured debt on financial distress among British households. 2005.

- [42] Andrew C Worthington. Debt as a source of financial stress in Australian households. *International Journal of Consumer Studies*, 30(1):2–15, 2006.
- [43] Brice Stone and Rosalinda Vasquez Maury. Indicators of personal financial debt using a multi-disciplinary behavioral model. *Journal of Economic Psychology*, 27(4):543–556, aug 2006.
- [44] Felipe Martínez, Rodrigo Cifuentes, Carlos Madeira, and Rubén Poblete-Cazenave. *Measurement of Household Financial Risk with the Survey of Household Finances*. Central Bank of Chile, 2013.
- [45] Marcelo Fuenzalida and Jaime Ruiz-Tagle. Riesgo financiero de los hogares. *Economía chilena*, 12(2):35–53, 2009.
- [46] Jaime Ruiz-Tagle and Francis Vella. Borrowing constraints and credit demand. *Documentos de Trabajo (Banco Central de Chile)*, (578):1, 2010.
- [47] Rodrigo Alfaro, Natalia Gallardo, and Roberto Stein. The determinants of household debt default. *Documentos de Trabajo (Banco Central de Chile)*, (574):1, 2010.
- [48] Jaime Ruiz-Tagle, Leidy García, and A Miranda. Proceso de endeudamiento y sobre endeudamiento de los hogares en Chile. *Documento de trabajo*, 703, 2013.
- [49] Yi Peng, Guoxun Wang, Gang Kou, and Yong Shi. An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing*, 11(2):2906–2915, mar 2011.
- [50] Edward I. Altman, Giancarlo Marco, and Franco Varetto. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance*, 18(3):505–529, may 1994.
- [51] Guoqiang Zhang, Michael Y. Hu, B Eddy Patuwo, and Daniel C. Indro. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1):16–32, jul 1999.
- [52] Jae H. Min and Young-Chan Lee. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4):603–614, may 2005.
- [53] D. Karthik Chandra, V. Ravi, and I. Bose. Failure prediction of dotcom companies using hybrid intelligent techniques. *Expert Systems with Applications*, 36(3):4830–4837, apr 2009.
- [54] Elena Fedorova, Evgenii Gilenko, and Sergey Dovzhenko. Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications*, 40(18):7285–7293, dec 2013.
- [55] Fatima Zahra Azayite and Said Achchab. Hybrid Discriminant Neural Networks for Bankruptcy Prediction and Risk Scoring. *Procedia Computer Science*, 83:670–674, jan 2016.

- [56] Nada Mselmi, Amine Lahiani, and Taher Hamza. Financial distress prediction: The case of French small and medium-sized firms. *International Review of Financial Analysis*, 50:67–80, mar 2017.
- [57] David Veganzones and Eric Séverin. An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems*, 112:111–124, aug 2018.
- [58] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.
- [59] Ying Mi. Imbalanced classification based on active learning SMOTE. *Research Journal of Applied Science Engineering and Technology*, 5:944–949, 2013.
- [60] William H Greene. The econometric approach to efficiency analysis. *The measurement of productive efficiency and productivity growth*, 1(1):92–250, 2008.
- [61] Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
- [62] Chris Bishop, Christopher M Bishop, and Others. *Neural networks for pattern recognition*. Oxford university press, 1995.
- [63] Guoqiang Zhang, B Eddy Patuwo, and Michael Y Hu. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62, 1998.
- [64] Casimir C Klimasauskas. Applying neural networks. Part 3: Training a neural network. *PC AI*, pages 20–24, 1991.
- [65] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533, 1986.
- [66] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [67] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification. 2003.
- [68] J-SR Jang. ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3):665–685, 1993.
- [69] Warren S Sarle. Neural Network FAQ, periodic posting to the Usenet newsgroup comp. ai. neural-nets. URL: <ftp://ftp.sas.com/pub/neural/FAQ.html>, 1997.
- [70] Timothy Masters. *Practical neural network recipes in C++*. Morgan Kaufmann, 1993.
- [71] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.