

# **ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE CIENCIAS**

**CONSTRUCCIÓN DE UN INDICADOR DE INCLUSIÓN FINANCIERA  
MULTIDIMENSIONAL Y SU RELACIÓN CON EL NIVEL DE  
POBREZA POR INGRESOS DE LOS INDIVIDUOS EN EL ECUADOR  
PERIODO 2018.**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERO EN CIENCIAS ECONÓMICAS Y FINANCIERAS**

**PROYECTO DE INVESTIGACIÓN**

**DANIELA CAROLINE ALBÁN MENDOZA**

daniela.alban@epn.edu.ec

**STALIN BENJAMIN BONILLA FRUTOS**

stalin.bonilla@epn.edu.ec

**DIRECTORA: ELIZABETH GUACHAMIN GUERRA, Ph.D**

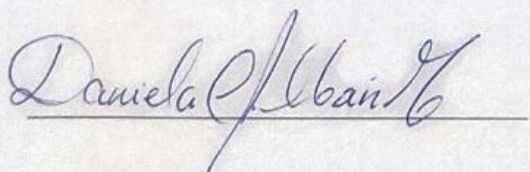
marcela.guachamin@epn.edu.ec

QUITO, MAYO 2023

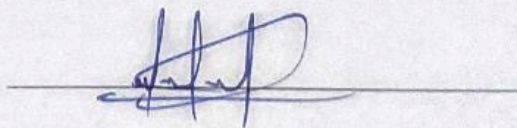
## DECLARACIÓN

Nosotros, Daniela Caroline Albán Mendoza y Stalin Benjamín Bonilla Frutos, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que hemos consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



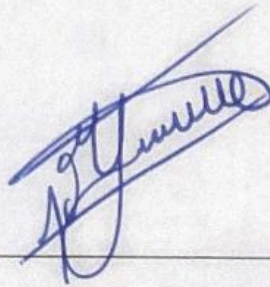
Daniela Caroline Albán Mendoza



Stalin Benjamín Bonilla Frutos

## CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Daniela Caroline Albán Mendoza y Stalin Benjamín Bonilla Frutos, bajo mi supervisión.

A handwritten signature in blue ink, appearing to read 'M. Guerra', is written over a horizontal line.

Marcela Elizabeth Guachamin Guerra, Phd.

## **AGRADECIMIENTO**

*A Dios, por darme fortaleza, salud y ser mi guía en cada paso.*

*A Juan Maximiliano, por llegar a mi vida y ser mi mayor motivación de ser mejor cada día.*

*A mis papás Anibal Maximiliano y Gilma Doris, por todo su amor, por ser mi principal ejemplo a seguir, por enseñarme a ser constante y perseverante, por todos los consejos, por ser mi apoyo fundamental a lo largo de mi vida y nunca dejarme sola.*

*A mi abuelita Melida y a mis hermanos Denise y Andrés quienes compartieron parte de su tiempo en el cuidado de mi más grande responsabilidad (Maximiliano) para que yo pueda culminar mis clases en la universidad.*

*A mis amigos de la universidad quienes hicieron de esta etapa una de las mejores de mi vida al ser un apoyo tanto dentro como fuera de las aulas. Principalmente a Stalin, mi “equipo dinamita”.*

*A mi tutora de tesis Dra. Marcela Guachamin por ser nuestra guía en este trabajo, por su exigencia, motivación y sobre todo por el tiempo compartido.*

*A todos mis profesores de la carrera quienes no sólo se limitaron a impartir materias, sino también que me han ayudado a mejorar constantemente.*

*A todas las personas que me brindaron su apoyo precisamente en los momentos cuando pensé que no podría “con todo a la vez”.*

***Daniela Caroline***

## **AGRADECIMIENTO**

*El concluir este trabajo me hace sentir muy agradecido con todas las personas que formaron parte de esta etapa tan importante en mi vida, y que de una forma u otra su apoyo y compañía me sirvió de fortaleza durante esta aventura.*

*Agradezco a mis padres, Neuton y Gladys, por ser las personas que más me apoyan en este mundo y que siempre han creído en mí. Les doy gracias por su compañía en cada uno mis pasos, por los sabios consejos que me han dado y me han servido como guía fundamental para superar las dificultades de la vida, por su interés en que yo sea mejor cada día y especialmente por todo el esfuerzo realizado para lograr que haya llegado hasta donde estoy hoy en día.*

*A mis hermanas, Allison y Belén, por ser la inspiración para cumplir mis objetivos y continuar cada día sirviéndoles de ejemplo. Les agradezco por ser por la luz de nuestra familia y motivarnos a superarnos cada día en cada cosa que hacemos.*

*A mis amigos, con quienes tuve el gusto de compartir este difícil camino desde sus inicios en nivelación hasta su finalización con este trabajo. Agradezco su ayuda y compañía, tanto en las aulas de clase como fuera de ellas, y que nos permitieron aprender mutuamente para crecer en lo académico, profesional y personal. Gracias por los bellos momentos vividos y que han dejado una huella en mí, siempre los recordaré con cariño.*

*A nuestra tutora, la Dra. Marcela Guachamin, por su colaboración y guía para la redacción de este trabajo y quien su conocimiento ha sido fundamental para culminar este trabajo.*

*A mí mismo, por no rendirme ante las adversidades de la vida y conseguir este logro. Todavía quedan metas por cumplir, sé fuerte y continúa cumpliendo tus sueños.*

**Stalin Benjamin**

## **DEDICATORIA**

*A Juan Maximiliano, la personita que estuvo a mi lado en todo este camino y ha sido mi principal motivación y fuerza.*

*A mis papás Anibal Maximiliano y Gilma Doris, mi apoyo incondicional.*

*Todos mis logros son por y para ustedes.*

*¡Los amo!*

***Daniela Caroline***

## **DEDICATORIA**

*Este trabajo está especialmente dedicado a toda mi familia a quien quiero mucho, ya que de no ser por ellos no hubiera obtenido este logro. Por su apoyo incondicional en cada momento desafiante en el que me he encontrado y por ser los pilares fundamentales en mi vida. Este es uno de muchos logros que vendrán para todos nosotros y de los sueños que cumpliremos.*

***Stalin Benjamin***

# Índice

<b>Índice de figuras</b> . . . . .	<b>IX</b>
<b>Índice de tablas</b> . . . . .	<b>X</b>
<b>Lista de anexos</b> . . . . .	<b>XI</b>
<b>Resumen</b> . . . . .	<b>XII</b>
<b>Abstract</b> . . . . .	<b>XIII</b>
<b>I Introducción</b> . . . . .	<b>1</b>
1.1 Antecedentes . . . . .	1
1.2 Objetivo General . . . . .	3
1.3 Objetivos Específicos . . . . .	3
<b>II Marco Teórico</b> . . . . .	<b>4</b>
2.1 Definición de Inclusión Financiera . . . . .	4
2.2 Pobreza por Ingresos . . . . .	6
2.3 Barreras a la Inclusión Financiera . . . . .	8
2.3.1 Barreras de Oferta . . . . .	8
2.3.2 Barreras de Demanda . . . . .	9
2.4 Inclusión Financiera y Pobreza . . . . .	11
2.5 Evidencia Empírica . . . . .	13
2.6 Medición Multidimensional de la Inclusión Financiera . . . . .	16
<b>III Datos y Metodología</b> . . . . .	<b>18</b>
3.1 Datos . . . . .	18
3.2 Estadística Descriptiva . . . . .	25
3.3 Metodología . . . . .	35
3.3.1 Modelos de Variable Latente . . . . .	35
3.3.2 Redes Neuronales Artificiales . . . . .	38
3.3.3 Modelos Variational Autoencoders (VAE) . . . . .	46
3.3.4 Validación Modelos VAE . . . . .	58
3.3.5 Modelos Logit . . . . .	63
3.3.6 Validación Modelos Logit . . . . .	66
<b>IV Resultados</b> . . . . .	<b>69</b>
4.1 Resultados de los Modelos Variational Autoencoder . . . . .	69
4.1.1 Modelos VAE: Subindicadores . . . . .	69
4.1.2 Modelo VAE: Indicador Multidimensional de Inclusión Financiera . . . . .	79



4.2	Análisis Cruzado del Indicador Multidimensional de Inclusión Financiera y Variables Sociodemográficas . . . . .	81
4.3	Resultados Logit . . . . .	92
4.3.1	Estimación . . . . .	92
4.3.2	Efectos Marginales . . . . .	97
<b>V</b>	<b>Conclusiones y Recomendaciones . . . . .</b>	<b>103</b>
<b>VI</b>	<b>Bibliografía . . . . .</b>	<b>116</b>
<b>VII</b>	<b>Anexos . . . . .</b>	<b>117</b>

# Índice de figuras

<b>Figura 2.1</b>	Línea de Pobreza y Pobreza Extrema por Ingresos Ecuador . . . . .	7
<b>Figura 3.1</b>	Acceso de Productos Financieros por Rango de Ingreso Per Cápita . . .	26
<b>Figura 3.2</b>	Acceso de Productos Financieros por Nivel de Educación . . . . .	27
<b>Figura 3.3</b>	Lejanía del Punto de Atención Financiero . . . . .	28
<b>Figura 3.4</b>	Frecuencia de Uso en Puntos de Atención Financieros . . . . .	29
<b>Figura 3.5</b>	Calificación de Puntos de Atención Financieros . . . . .	30
<b>Figura 3.6</b>	Desinterés Productos Financieros . . . . .	32
<b>Figura 3.7</b>	Ahorro de Dinero . . . . .	33
<b>Figura 3.8</b>	Estructura Básica de una Red Neuronal . . . . .	39
<b>Figura 3.9</b>	Algoritmo Ejecutado por una Neurona . . . . .	41
<b>Figura 3.10</b>	Funciones de Activación Utilizadas . . . . .	42
<b>Figura 3.11</b>	Proceso del Algoritmo Backpropagation . . . . .	45
<b>Figura 3.12</b>	Estructura General Redes Neuronales VAE . . . . .	55
<b>Figura 4.1</b>	Valor de Shapley: Modelos VAE . . . . .	74
<b>Figura 4.2</b>	Estructura de Red Neuronal Modelo VAE Acceso . . . . .	77
<b>Figura 4.3</b>	Estructura de Red Neuronal Modelo VAE Bienestar . . . . .	77
<b>Figura 4.4</b>	Estructura de Red Neuronal Modelo VAE Uso . . . . .	78
<b>Figura 4.5</b>	Estructura de Red Neuronal Modelo VAE Calidad . . . . .	78
<b>Figura 4.6</b>	Red Neuronal VAE: Inclusión Financiera Multidimensional . . . . .	81
<b>Figura 4.7</b>	Distribución Indicador Multidimensional Inclusión Financiera . . . . .	82
<b>Figura 4.8</b>	Indicador Multidimensional: Dispersión y Tendencia por Edad . . . . .	88
<b>Figura 4.9</b>	Indicador Multidimensional: Promedio por Categoría de Ocupación . . .	89
<b>Figura 4.10</b>	Indicador Multidimensional: Boxplot por Nivel de Educación . . . . .	90
<b>Figura 4.11</b>	Indicador Multidimensional: Histograma de Densidad por Ingreso Per Capita . . . . .	91

# Índice de tablas

<b>Tabla 3.1</b>	Lista de Variables Usadas en la Creación del Indicador Multidimensional	20
<b>Tabla 3.2</b>	Lista de Variables Modelos Logit	23
<b>Tabla 3.3</b>	Acceso de Productos Financieros por Área	25
<b>Tabla 3.4</b>	Deseo de Productos Financieros	31
<b>Tabla 3.5</b>	Aprobación de Crédito por Rango de Ingreso Per Cápita	33
<b>Tabla 3.6</b>	Ahorro de Dinero por Área	34
<b>Tabla 3.7</b>	Ahorro de Dinero por Sexo	34
<b>Tabla 4.1</b>	Configuración VAE Acceso y Resultados de Pérdida	70
<b>Tabla 4.2</b>	Configuración VAE Bienestar y Resultados de Pérdida	71
<b>Tabla 4.3</b>	Configuración VAE Uso y Resultados de Pérdida	72
<b>Tabla 4.4</b>	Configuración VAE Calidad y Resultados de Pérdida	73
<b>Tabla 4.5</b>	Simulaciones Modelos VAE: Subindicadores	76
<b>Tabla 4.6</b>	Simulaciones Modelos VAE: Indicador Multidimensional	79
<b>Tabla 4.7</b>	Simulaciones Modelos VAE: Indicador Multidimensional	80
<b>Tabla 4.8</b>	Test de Multimodalidad de Hartigan	83
<b>Tabla 4.9</b>	Indicador Promedio por Área: Región Costa	84
<b>Tabla 4.10</b>	Indicador Promedio por Área: Región Sierra	85
<b>Tabla 4.11</b>	Indicador Promedio por Área: Región Oriente	86
<b>Tabla 4.12</b>	Indicador Promedio por Rango de Edad	87
<b>Tabla 4.13</b>	Resultado de las Estimaciones Logit de los Determinantes de la Inclusión Financiera en Ecuador	93
<b>Tabla 4.14</b>	Efectos Marginales de las Estimaciones Logit de los Determinantes de la Inclusión Financiera en Ecuador	97

## **Lista de anexos**

<b>ANEXO A.</b> Datos Perdidos . . . . .	117
<b>ANEXO B.</b> Validación Modelos VAE . . . . .	119
<b>ANEXO C.</b> Definición Dimensiones Inclusión Financiera . . . . .	120
<b>ANEXO D.</b> Validación de Modelos Logit . . . . .	121
<b>ANEXO E.</b> Categorización Indicador Multidimensional de Inclusión Financiera . . . . .	125

## Resumen

La inclusión financiera ha tomado relevancia a nivel mundial debido a que ha demostrado ser una herramienta que permite a los individuos mejorar su calidad de vida en el sentido de que pueden satisfacer sus necesidades y hacer frente a situaciones adversas mediante el acceso y uso de productos y servicios financieros de calidad. Sin embargo, su medición ha resultado compleja debido a su concepto multidimensional. Por ello, la presente investigación utiliza la base de datos ENEMDU de diciembre del 2018 y la metodología de Redes Neuronales *Variational Auto-encoder* para desarrollar un índice de inclusión financiera que captura las cuatro dimensiones que la definen: acceso, uso, calidad y bienestar. Posteriormente, mediante un modelo *logit* se evalúa las características socio-demográficas, de situación económica y financiera que determinan el nivel de inclusión de los individuos en Ecuador. Principalmente, se busca analizar la relación que existe con la pobreza por ingresos. Los resultados del índice mostraron que el nivel de inclusión financiera en Ecuador es alto únicamente para una pequeña parte de los individuos de la muestra. Posteriormente, al categorizar el indicador entre incluido y no incluido financieramente, se encontró que los principales factores que disminuyen la probabilidad de encontrarse incluido financieramente son los bajos niveles de educación y encontrarse en situación de pobreza.

**Palabras Clave:** Inclusión Financiera, Pobreza, Variational Autoencoder, índice multidimensional, logit.

## Abstract

Financial inclusion has gained global relevance due to its effectiveness in enhancing the quality of life for individuals, enabling them to fulfill their needs and cope with unexpected situations through access to quality financial products and services. However, its measurement has been challenging due to its multidimensional nature. Thus, this study employs the ENEMDU database from December 2018 and the *Variational Autoencoder* Neural Networks methodology to construct a financial inclusion index that encompasses the four dimensions of access, use, quality, and well-being. Subsequently, a logit model is utilized to examine the socio-demographic, economic, and financial characteristics that determine individuals' level of inclusion in Ecuador and analyze the relationship with income poverty. The index's results demonstrate that the level of financial inclusion in Ecuador is high for only a small portion of the sample. Upon categorizing the indicator between financially included and not included, the analysis revealed that low levels of education and poverty are the main factors that decrease the likelihood of being financially included.

**Palabras Clave:** Financial Inclusion, Poverty, Variational Autoencoder, Multi-dimensional Index, logit.

# Capítulo I

## Introducción

### 1.1. Antecedentes

El Banco Mundial (2018) menciona que la inclusión financiera permite a las personas y empresas tener acceso a servicios y productos financieros, los cuales sirven como mecanismos para satisfacer necesidades tales como realizar transacciones, efectuar pagos, ahorrar, acceder a créditos y obtener seguros.

Por otro lado, la Alianza para la Inclusión Financiera (2010) menciona que la inclusión financiera está compuesta por cuatro pilares fundamentales: acceso, uso, calidad y bienestar. En este sentido, de acuerdo con Cano et al. (2014), el acceso a los servicios financieros es una condición necesaria pero no suficiente para lograr la inclusión financiera, ya que además es fundamental garantizar el uso efectivo de los mismos. En tal sentido, el fomentar la inclusión financiera también implica brindar servicios de calidad, de modo que se facilite el acceso y uso de servicios financieros, generando así bienestar hacia el beneficiario (Cull et al., 2014). Por lo tanto, al tratarse de un concepto multidimensional, la medición de la inclusión financiera resulta compleja.

Sotomayor et al. (2018) y Narváez et al. (2020) explican, mediante modelos probit, que

la inclusión financiera en Perú y Colombia se enfoca en el acceso a crédito en una entidad formal. Así mismo, Rodríguez-Raga y Rodríguez (2016); Zins y Weill (2016) y Fungáčová y Weill (2015) evalúan si un hogar se encuentra incluido financieramente de acuerdo al acceso que posean a al menos un producto financiero formal.

En este contexto, la inclusión financiera debe evaluar las características multidimensionales que identifican al acceso, uso, bienestar, calidad y la posible autoexclusión (Rao et al., 2017). Cámara y Tuesta (2014) plantearon una metodología para la creación de un indicador multidimensional con base en las dimensiones de acceso, uso y barreras reflejadas en ochenta y dos países.

Con estos antecedentes, esta investigación busca desarrollar un índice de inclusión financiera multidimensional compuesto de los cuatro elementos mencionados por la teoría. Para el efecto, se utiliza la información obtenida en el módulo financiero por la encuesta ENEMDU en diciembre del año 2018, utilizando la metodología de redes neuronales *Variational Autoencoder* (VAE) (Kingma y Welling, 2013).

El indicador de inclusión financiera propuesto identificará el nivel de inclusión del segmento poblacional de interés en el sistema financiero formal, considerando las cuatro dimensiones: acceso, uso, calidad y bienestar. Además, se caracterizará a las personas que están incluidas o no financieramente de acuerdo con parámetros sociodemográficos, financieros y considerando su nivel de pobreza por ingresos. Los hallazgos obtenidos buscan aportar al diseño de políticas, programas y estrategias para el desarrollo económico mediante el enfoque de la inclusión financiera.



## 1.2. Objetivo General

Construir un índice de inclusión financiera, como indicador para evaluar su relación con la pobreza por ingresos en el Ecuador, durante el año 2018.

## 1.3. Objetivos Específicos

- Construir un índice de inclusión financiera con base en 4 dimensiones: acceso, uso, calidad y bienestar; estimado mediante la metodología de redes neuronales *Variational Auto-encoder* (VAE).
- Analizar la relación existente entre inclusión financiera y pobreza en el Ecuador durante el año 2018.
- Identificar cómo influye la pobreza en la inclusión financiera a través del análisis de características sociodemográficas, económicas y financieras de los jefes de hogar en Ecuador año 2018, utilizando un modelo logit binomial.

# Capítulo II

## Marco Teórico

### 2.1. Definición de Inclusión Financiera

La inclusión financiera ha alcanzado una relevancia significativa en los últimos años por el incremento del acceso y uso de productos financieros a nivel mundial, sin embargo, no existe un consenso sobre su definición. Chakrabarty (2011) define a la inclusión financiera como el proceso de garantizar que todos los sectores de la sociedad, incluidos los grupos vulnerables, como los grupos de bajos ingresos, tengan acceso a productos y servicios financieros a costos asequibles y transparentes por parte de las instituciones financieras. Por otro lado, Sarma (2008) agrega que la inclusión financiera, además de definirse como un proceso que garantiza el acceso al sistema financiero y su disponibilidad, también debe incluir el uso efectivo del mismo, cuya finalidad es la construcción de un sistema financiero inclusivo. Asimismo, Arellano et al. (2014) la definen como el proceso donde se maximiza el acceso y uso de los servicios financieros formales y se minimizan las barreras involuntarias percibidas por los individuos que no participan en el sistema financiero formal, para lo cual los servicios financieros deberían ser de calidad y los individuos deberían obtener bienestar a través de éstos.

Organismos internacionales, como el CGAP (2011) y el Banco Mundial (2022), señalan

que la inclusión financiera significa que todos los adultos en edad de trabajar puedan acceder a productos y servicios financieros útiles que atiendan sus necesidades, tales como ahorro, crédito, pagos, transacciones y seguros; los cuales deben ser provistos de manera responsable y sostenible, de tal manera que las personas utilicen el sistema financiero formal y no opciones informales.

De acuerdo al programa propuesto por el Banco Central del Ecuador titulado “Estrategia Nacional de Inclusión Financiera 2020-2024”, la inclusión financiera se refiere a los individuos y empresas que pueden acceder a utilizar productos y servicios financieros de calidad, y lo puedan hacer de una manera informada. Además, estos productos y servicios deben ser ofrecidos por las entidades financieras formales de forma transparente, responsable, sostenible y deben responder a las necesidades de la población.

Uno de los organismos comprometidos con la medición de la inclusión financiera es la Alianza para la Inclusión Financiera (AFI), la cual en su documento denominado “La medición de la inclusión financiera para entes reguladores: Diseño e implementación de encuestas” menciona que la inclusión financiera en sí misma es un concepto multidimensional que abarca cuatro elementos:

- I. **Acceso:** Se refiere a las capacidades existentes que permiten utilizar los productos y servicios financieros ofrecidos por las instituciones formales, como precios asequibles o la proximidad física a los puntos de atención financieros.
- II. **Calidad:** Indica la relevancia del producto o servicio financiero dentro de las necesidades del estilo de vida de los consumidores. Esta dimensión abarca la experiencia de los consumidores manifestada a través de actitudes y opiniones hacia aquellos productos que en la actualidad se encuentran disponibles.
- III. **Uso:** Esta dimensión abarca la adopción de servicios bancarios, pone mayor énfasis en la

regularidad, frecuencia y duración de su utilización a lo largo del tiempo de los productos y servicios financieros.

IV. **Bienestar:** Es el impacto que un dispositivo o servicio financiero ha tenido en la vida de los consumidores, entre otros: cambios en el consumo, las actividades empresariales, educación financiera y el bienestar general.

La inclusión financiera es entendida bajo diferentes conceptos; sin embargo, para el presente estudio se adopta la definición de la Alianza para la Inclusión Financiera.

## **2.2. Pobreza por Ingresos**

La pobreza, al igual que muchos otros conceptos que han surgido en las ciencias sociales, carece de una definición única y universal. Es así que, abarca una multiplicidad de elementos de diversa índole de la vida de las personas y, por tanto, es casi imposible resumirla en una sola definición. Aunque existen distintos enfoques por los cuales se puede entender a la pobreza, en términos generales, ha sido comprendida como la situación en la cual las personas carecen de recursos necesarios para alcanzar un nivel de bienestar compatible con la dignidad humana (CEPAL, 2018).

Para comprender la pobreza, en este estudio se utiliza el enfoque de pobreza por ingresos, el cual considera que una persona está en situación de pobreza cuando no cuenta con los ingresos suficientes para satisfacer sus necesidades básicas. Este enfoque mide la pobreza comparando el ingreso per cápita del individuo con una línea de pobreza establecida, si su nivel de ingreso es inferior a dicha línea se considera que se encuentra en situación de pobreza (Banco Mundial , 2017).

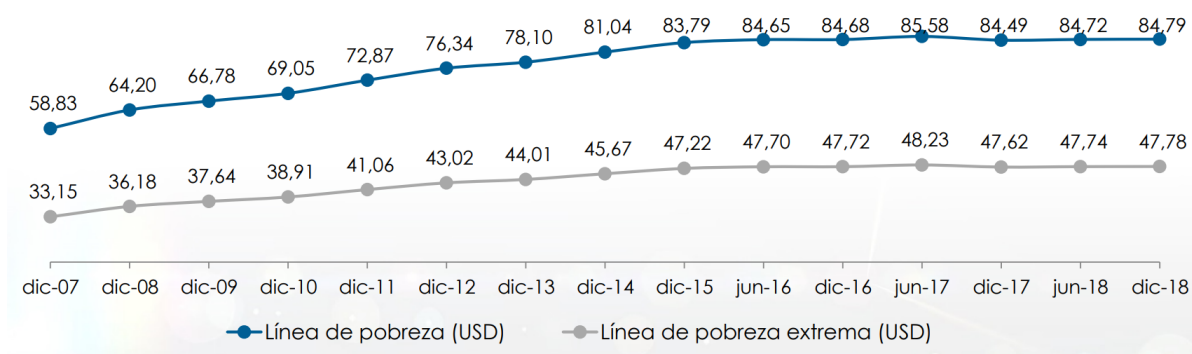
Según la CEPAL (2018), existen dos formas para determinar la línea de pobreza. La primera, consiste en establecer una línea de pobreza con base al costo de una canasta básica de

alimentos, al que se le agrega un monto destinado a cubrir otro tipo de necesidades no alimentarias. Mientras que, la segunda forma consiste en tomar un valor derivado de un punto de la función de distribución de ingresos, como es el 60 % de la mediana del ingreso de los hogares del país.

En el caso de Ecuador, el Instituto Nacional de Estadísticas y Censos obtiene la línea de pobreza utilizando el primer método antes mencionado, además agrega la línea de pobreza extrema, la cual considera únicamente el componente alimentario de la canasta básica. A partir del 2017 esta línea se calcula actualizando la línea oficial de pobreza por consumo mediante el Índice de Precios al Consumidor (IPC)<sup>1</sup>.

La Figura 2.1 expone la evolución nominal de las líneas de pobreza y pobreza extrema, la cual muestra que en diciembre de 2018 la línea de pobreza se estableció en USD 84,79 mensuales per cápita, mientras que la línea de pobreza extrema se ubicó en USD 47,78 mensuales per cápita. Por lo tanto, aquellos individuos con un ingreso total per cápita inferior a este valor fueron considerados como pobres por ingresos.

**Figura 2.1:** Línea de Pobreza y Pobreza Extrema por Ingresos Ecuador



*Fuente:* Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) - INEC

*Nota :* Los datos se encuentran expresados en dólares mensuales corrientes.

<sup>1</sup>La línea de pobreza por consumo se obtuvo a partir de la Encuesta de Condiciones de Vida 5ª ronda.

## **2.3. Barreras a la Inclusión Financiera**

La falta de inclusión financiera se presenta por una serie de factores tanto del lado de la oferta como de la demanda, afectando así a varios segmentos de la población, principalmente a los estratos más pobres, ya que al restringirles la movilidad del capital se los incapacita de ser partícipes del proceso económico (Devlin, 2005). A continuación, se explica las barreras de oferta y de demanda que están relacionadas con la inclusión financiera.

### **2.3.1. Barreras de Oferta**

E. Pérez y Titelman (2018) señalan que, en general, el análisis de la inclusión financiera, tanto teórico como empírico, se ha centrado casi exclusivamente en el lado de la oferta. Los autores argumentan que la falta de inclusión financiera se debe a barreras de entrada al sistema financiero, refiriéndose específicamente a barreras de precios, ya que las instituciones financieras ofrecen financiamiento a precios (tasas de interés) demasiado altos y en condiciones que son demasiado restrictivas. De acuerdo con la Alianza para la Inclusión Financiera (2010), estos costos elevados generan la exclusión de aquella población vulnerable de escasos recursos, los cuales tienen ingresos irregulares o no cuentan con un empleo estable.

Por otra parte, Roa (2013) menciona que también son los costos elevados de transacción los que reducen la oferta de servicios financieros, sobre todo a poblaciones de zonas rurales, debido a que no les resulta rentable abrir un punto de prestación de servicios financieros en estas zonas. Álvarez-Gamboa et al. (2021) describen a esta barrera como restricción de accesibilidad física por parte del sistema financiero, misma que incluye los costos relacionados con las actividades de transacción entre las instituciones financieras y el demandante referente a la estructura física. Con respecto a esto, Duncombe y Boateng (2009) han señalado que el desarrollo tecnológico facilitaría el acceso al reducir la necesidad de viajar largas distancias.

Según Álvarez-Gamboa et al. (2021) otra barrera a la inclusión financiera es la elegibilidad, la cual implica la escasa información entre ofertante y demandante de los productos y servicios financieros. En particular, se trata de los costos que surgen de la asimetría de información que produce el sistema financiero. Al respecto, según E. Pérez y Titelman (2018), la asimetría de información es la falla de mercado que afecta a la inclusión financiera, produciendo dos consecuencias importantes dentro del mercado crediticio. La primera es la selección adversa: los individuos que más desean solicitar créditos son aquellos que precisamente son los menos atractivos desde el punto de vista del prestamista al ser los más riesgosos. Mientras que la segunda es el riesgo moral: el riesgo de que el deudor tome una acción que le cause daño al prestamista, por ejemplo, que utilice los fondos prestados para consumo y que posteriormente no pueda hacer frente a las obligaciones de la deuda adquirida.

### **2.3.2. Barreras de Demanda**

Con respecto al lado de la demanda, autores como Beck y De la Torre (2007), Beck y Demirgüç-Kunt (2008) resaltan que cuando la decisión de no utilizar los productos y servicios financieros está dada por preferencias intertemporales y oportunidades de inversión por parte de los individuos, no representa un problema. Sin embargo, dados ciertos factores, los individuos a pesar de tener acceso podrían no hacer uso de estos productos y servicios financieros. Como ejemplo de estos factores tenemos: desconocimiento o falta de educación financiera, falta de ingresos o empleo, desconfianza en las instituciones financieras, miedo al endeudamiento o autoexclusión como resultado psicológico de una discriminación sistemática en el pasado. En cualquiera de los casos, el acontecimiento de estos factores representaría un problema de inclusión financiera.

De acuerdo con Álvarez-Gamboa et al. (2021), el desconocimiento o la falta de educación financiera provoca el uso ineficiente, así como también la exclusión misma de los productos

y servicios financieros. A su vez, Dassatti y Marino (2017) precisan que los agentes económicos que tienen una menor educación financiera, son aquellos que se encuentran en situación de pobreza. Por otra parte, Raccanello y Guzmán (2014) y Roa y Carvallo (2018) mencionan la importancia de que los individuos tengan conocimiento sobre los distintos productos y servicios que ofrecen las entidades financieras formales. La razón se debe a que de esta forma la inclusión financiera les generará un mayor bienestar y podrán tomar decisiones financieras óptimas, además de lograr administrar conscientemente su capital en actividades que les resulten beneficiosas.

Según E. Pérez y Titelman (2018) la falta de educación financiera también podría influir en la desconfianza o percepción que los agentes tienen sobre el sistema financiero. Por su parte, Roa (2010) menciona que esta desconfianza en las entidades financieras se ha generado debido a las crisis económicas que se han desarrollado a lo largo de la historia, así como también a la asimetría de información y especulaciones. Razones las cuales han provocado que los usuarios no se encuentren motivados a adherirse a los servicios y productos financieros. En este sentido, Roa (2010) menciona que las instituciones financieras deben funcionar como intermediarios de calidad para así dinamizar los recursos, llegar a los sectores más necesitados y reducir las brechas de pobreza y desigualdad.

Al considerar otras barreras de demanda, son los costos de información y transacciones los motivos que limitan el acceso a la población que se encuentra en situación de pobreza. Asimismo, la falta de garantías y no tener historial crediticio restringe el acceso a crédito a este segmento de la población, por lo cual se intensifica la desigualdad de ingresos al impedir el flujo de capital a los individuos pobres con altas expectativas de inversión. Por lo tanto, la flexibilidad de las restricciones crediticias beneficiaría acertadamente a los pobres (Álvarez-Gamboa et al., 2021).

En términos generales, Beck y De la Torre (2007) sugieren que el precio y el ingreso son los



principales factores que influyen en la demanda de los servicios financieros, tales como pagos y ahorros. Esto debido al hecho de que, el desarrollo económico y el incremento asociado en el ingreso per cápita, aumenta la necesidad de servicios más sofisticados. Asimismo, autores como Allen et al. (2016) han encontrado que la principal barrera por el lado de la demanda es la insuficiencia de ingresos.

Por otra parte, Claessens (2006) y Roa y Carvalho (2018) mencionan que la demanda no está incentivada únicamente por factores económicos, sino que también se puede dar por factores socioculturales como por ejemplo la religión, lo que llevaría a que los individuos se excluyan voluntariamente del sistema financiero.

## **2.4. Inclusión Financiera y Pobreza**

En la sección anterior se presentaron las barreras existentes a la inclusión financiera, donde se pudo notar que el nivel de ingresos de los individuos es un factor principal que influye en la inclusión financiera. Esto debido a que las personas que se encuentran en los quintiles más bajos de ingresos tienen mayores dificultades para acceder al sistema financiero formal. A tal efecto, esta sección profundiza el tema y realiza una recopilación de literatura sobre la relación existente entre la inclusión financiera y la pobreza por ingresos.

La literatura existente referente a la relación entre pobreza e inclusión financiera es amplia, varios estudios analizan el enfoque de cómo la inclusión financiera incide en la pobreza. Así, por ejemplo, Beck et al. (2009) y Sehrawat y Giri (2016) han mostrado que la inclusión financiera es una herramienta que genera un impacto positivo en la economía, ya que reduce la pobreza indirectamente al lograr aumentar el crecimiento económico. Sin embargo, de acuerdo con autores como Greenwood y Jovanovic (1990) y Narváez et al. (2020), este es un efecto que se genera a largo plazo, mientras que a corto plazo el desarrollo financiero no es óptimo debido a que acentúa la desigualdad, ya que son precisamente las personas pobres, quienes son

excluidas del sistema financiero formal.

Martínez (2008) señala que las instituciones financieras se centran en trabajar con grupos objetivos con los cuales se pueda recuperar los fondos y no con aquellos que se encuentran en situación de pobreza, por lo tanto, el ser parte de esta población disminuiría la posibilidad de ser incluido financieramente. En este sentido, García (2016) menciona que existe una tendencia a que los pobres no puedan acceder plenamente a los servicios de las entidades financieras y, en consecuencia, el desarrollo del sistema financiero no les resulta beneficioso.

La evidencia empírica ha demostrado que el nivel de ingresos o a su vez, la pobreza es un factor que influye en el hecho de que un individuo se encuentre incluido o no financieramente. Así, por ejemplo, Demirguc-Kunt et al. (2013) utilizando la base de datos Global Findex 2011, encontraron que en los países en desarrollo los individuos pertenecientes a los quintiles de menor ingreso, tienen menor probabilidad de tener acceso a los productos y servicios brindados por entidades financieras formales con respecto a los individuos que se encuentran en el quintil más alto de ingresos. Este resultado fue coincidente con los obtenidos por autores como, Cámara et al. (2013), en Perú; Hoyo et al. (2013), en México y Tuesta et al. (2015), para el caso de Argentina.

En este sentido, Narváez et al. (2020) se centraron en probar la incidencia de la pobreza monetaria en la inclusión financiera de los hogares colombianos. Para este objetivo hicieron uso de un modelo *probit*, en donde el indicador de inclusión financiera tomaba el valor de 1 si el jefe de hogar había tenido acceso a crédito en una entidad financiera formal. Sus resultados sugirieron que existe una relación inversa entre pobreza monetaria e inclusión financiera. Por su parte, Rodríguez-Raga y Rodríguez (2016) encontraron que en Colombia, aquellos individuos con mayor nivel de ingreso tienen mayor probabilidad de acceder a productos de ahorro, crédito y seguros.

De acuerdo con Narváez et al. (2020), la evidencia existente muestra que el principal desafío

al que se enfrenta la inclusión financiera es lograr llegar a los segmentos más pobres de la población, puesto que es precisamente por su condición de pobreza que han sido excluidos del sistema financiero. Según Maron (2014) esto se debe a la configuración económica de los estratos sociales, en donde las personas con ingresos estables y que poseen patrimonio, son consideradas aptas para ser partícipes de los servicios financieros, perpetuando así la restricción de los servicios financieros a los grupos más pobres.

## **2.5. Evidencia Empírica**

En esta sección se presenta la evidencia empírica relacionada con las características socio-demográficas que determinan la inclusión financiera y las distintas formas en que los autores la miden.

Varios autores han analizado los determinantes de inclusión financiera, midiéndola como la tenencia o uso de productos y servicios en instituciones financieras formales, tales como: cuentas, ahorros y créditos. Así, por ejemplo, Zins y Weill (2016) en África y Fungáčová y Weill (2015) en China, utilizando los datos de la encuesta Global Findex 2011 y 2012, respectivamente, estimaron tres modelos *probit* donde las variables dependientes eran cada uno de los indicadores (uso de cuentas, ahorros y crédito). En ambos casos encontraron que la edad tiene una relación no lineal con los tres indicadores. Según Fungáčová y Weill (2015) esto puede ser explicado en términos de un “efecto generacional” que pudo derivar del lado de la demanda, debido a que las personas mayores suelen ser más reticentes a emplear los servicios financieros formales, mientras que, por el lado de la oferta, explican que las instituciones financieras podrían poner menor esfuerzo en atraer clientes de mayor edad. Por otra parte, Zins y Weill (2016) encontraron que el ser mujer disminuye significativamente la probabilidad de tener una cuenta o ahorrar en una institución financiera, sin embargo, no encontraron resultados significativos en la tenencia de crédito. Finalmente, concluyeron que particularmente las variables de educación

e ingresos favorecen al acceso de cuentas, ahorros y créditos. En contraposición, Fungáčová y Weill (2015)) encontraron que el nivel de educación no tiene ningún efecto sobre el ahorro formal y explicaron que este resultado podría ser debido a que en China existe una población relativamente alta que cuenta con ahorros. Por otra parte, la variable de ingreso solo resultó ser significativa en el modelo de tenencia de cuenta, los individuos que pertenecen a los quintiles más bajos tienen menor probabilidad de acceder a este producto con respecto al quintil más alto de ingreso, pero que en el caso de ahorro el ingreso no resultó ser significativo. Mientras tanto, en lo que respecta al acceso al crédito, solo el cuarto quintil de ingresos resultó ser negativo y estadísticamente significativo, es decir, solo el 20% de las personas justo debajo del 20% más rico tiene un uso significativamente menor del crédito formal con respecto a los más ricos.

Demirgüç-Kunt et al. (2013) se centraron en analizar las diferencias de género en el uso de cuentas, ahorros y créditos en instituciones financieras formales utilizando datos a nivel individual de 98 países en desarrollo en 2011. Para el caso de cuentas, estimaron un modelo *probit* de respuesta binaria, mientras que, para el caso de ahorro y crédito, estimaron modelos *logit* multinomial. En el segundo caso, la variable dependiente fue clasificada en: 'ninguno', 'ahorro o crédito formal' y 'ahorro o crédito informal'. Los resultados indicaron la existencia de una brecha de género en los países en desarrollo. Incluso después de controlar características individuales como ingresos, educación, situación laboral, residencia en zona rural y la edad, el ser mujer sigue estando significativamente relacionado con el uso de los servicios financieros.

Cámara et al. (2013), utilizando la Encuesta Nacional de Hogares 2011 en Perú y un modelo *probit* con la variable dependiente que considera el uso de productos financieros formales, identificaron que los grupos que han sido tradicionalmente más vulnerables son aquellos que se encuentran excluidos del sistema financiero formal: las mujeres, los jóvenes y las personas que viven en zonas rurales.

Hoyo et al. (2013) exploran los factores que determinan la inclusión financiera en México

utilizando los datos de la Encuesta Nacional de Inclusión Financiera 2012. Para este propósito, primero se construyeron indicadores de inclusión financiera utilizando el método de análisis de correspondencias múltiples. Se tomaron en cuenta la tenencia de productos de crédito y ahorro, tanto de forma conjunta (Indicador Agregado) como individualmente (Indicador de Ahorro e Indicador de Crédito). Es así que, los autores encontraron que la educación es un factor clave para explicar la inclusión financiera, además, el ser jefe de hogar, estar casado o en unión libre se relacionaron positivamente con los tres indicadores. A la vez, el ser mujer, no resultó ser significativo en el modelo con el indicador agregado, pero sí en los indicadores individuales. Así también, se encontró que para productos de crédito la relación es positiva, mientras que para productos de ahorro la relación es negativa, lo que implica que tienen más exclusión desde el punto de vista de productos financieros como cuentas, depósitos o inversiones.

Tuesta et al. (2015) realizaron un análisis de la inclusión financiera para el caso de Argentina utilizando datos del Global Findex 2012, donde adoptaron tres dimensiones: acceso, calidad y uso de los servicios y productos financieros. Señalaron que el acceso a los servicios formales se basa en los canales tradicionales: sucursales y ATM's; en cuanto a calidad, encontraron que la edad y los ingresos son los factores que incidieron en la percepción de las diferentes barreras de exclusión involuntaria. Finalmente, para la dimensión de uso estimaron cinco modelos *probit* en donde cada uno tenía como variable dependiente a la tenencia de los siguientes productos o servicios financieros: cuenta en una institución financiera formal, tenencia de tarjeta de débito, tarjeta de crédito, pagos electrónicos y crédito formal. Los resultados mostraron que la edad, el nivel de educación y los ingresos son factores que determinan la tenencia de los productos o servicios antes mencionados.

Trivelli (1997) señala que el ahorro formal es una variable relevante para la inclusión financiera, puesto que, un mayor nivel de ahorro da apertura a que las familias tenga acceso a otros productos financieros. Por otra parte, Cano et al. (2014); Devlin (2005); Rhine y Greene (2006)

incluyeron el tamaño del hogar como una característica que determina la inclusión financiera y encontraron que, en general, mientras más grande es el tamaño del hogar, menor es el nivel de inclusión financiera.

En general, la literatura ha mostrado que entre los principales determinantes que favorecen al acceso de los productos financieros son tener un mayor nivel de ingresos (Cano et al., 2014; Pabón et al., 2006; Trivelli, 1997) y un mayor nivel de educación (Al-Hussainy et al., 2008; Cano et al., 2014; Johnson y Nino-Zarazua, 2011; Kumar, 2005; Pabón et al., 2006). Asimismo, encontrarse en pleno empleo ha sido una característica que de acuerdo con Kumar (2005) favorece principalmente al acceso a créditos. Algunos autores como Al-Hussainy et al. (2008); Cano et al. (2014); Devlin (2005) concluyeron que el estado civil casado también favorece al acceso de los productos financieros.

## **2.6. Medición Multidimensional de la Inclusión Financiera**

A lo largo de los años, varios estudios han desarrollado diferentes metodologías con el objetivo de crear una medida integral que abarque la naturaleza multidimensional de la inclusión financiera. En esta sección, se explorará la literatura existente en torno a la medición de la inclusión financiera.

Sarma (2008) propone un índice multidimensional basado en la metodología usada para desarrollar el índice de desarrollo humano (IDH) y usando tres dimensiones de inclusión financiera. El indicador fue creado en dos fases utilizando la definición de distancia euclidiana normalizada. La primera fase consistió en crear subindicadores para cada dimensión, mientras que en la segunda fase fueron agregados en una única medida. Park y Mercado (2015), siguiendo un enfoque similar, desarrollaron un indicador multidimensional esta vez usando cinco dimensiones. Sin embargo, este enfoque asignó pesos arbitrarios a cada dimensión, lo que podría sesgar la medición correcta del indicador multidimensional (Levine, 2005).

Amidžic et al. (2014) proponen una metodología más robusta mediante la técnica de análisis factorial y usando cinco dimensiones, las cuales fueron agregadas usando pesos estadísticos siguiendo una media geométrica ponderada. Por otro lado, Cámara y Tuesta (2014) proponen una metodología basada en el análisis de componentes principales en dos etapas y usando variables de acceso, uso y barreras. La primera fase consistió en crear subindicadores de cada dimensión, los cuales son agregados en la segunda fase en un único indicador multidimensional. En este enfoque se asignan pesos a cada dimensión de inclusión financiera, mismos que son estimados mediante las ecuaciones de la metodología. Por lo tanto, se excluye la subjetividad de darle un peso a cada dimensión de inclusión financiera.

En general, la revisión de la literatura discutida previamente muestra que han existido esfuerzos para el desarrollo de una medida multidimensional de inclusión financiera. Los indicadores de estos estudios se basan principalmente en dos enfoques: paramétrico y no paramétricos. Como ejemplo del primer enfoque se tienen los estudios de Sarma (2008) y Park y Mercado (2015) dado que asignan la importancia de cada indicador eligiendo sus pesos de forma exógena. Mientras tanto, los estudios de Amidžic et al. (2014) y Cámara y Tuesta (2014) son ejemplos del segundo enfoque, ya que asignan la importancia de los indicadores de forma endógena a partir de la estructura de los datos. En tal sentido, métodos estadísticos como el análisis de componentes principales son usualmente preferidos dado que permiten obtener un resultado haciendo la menor cantidad posible de supuestos (Freund y Rubin, 1980).

# Capítulo III

## Datos y Metodología

Este capítulo describe la fuente de datos utilizada en la construcción del indicador de inclusión financiera, así como las metodologías empleadas en el estudio. Además, se presenta un análisis descriptivo de los datos utilizados en este estudio.

### 3.1. Datos

En la presente investigación se utilizaron los datos provenientes de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) correspondiente a diciembre del año 2018. La información empleada para la construcción del índice multidimensional fue obtenida del módulo financiero, el cual contiene información sobre el uso y acceso de productos financieros. Así mismo, la información financiera de la encuesta fue combinada con los datos socio-demográficos de los jefes de hogar (INEC, 2018a). De este modo, la base de datos utilizada en este estudio contiene 6926 individuos.

La Encuesta Nacional Empleo, Desempleo y Subempleo (ENEMDU) la realiza el Instituto Nacional de Estadística y Censos y sirve como un instrumento para determinar información estadística sobre la situación del empleo, actividades económicas y fuentes de ingresos de los



individuos del país. La encuesta se basa en un muestreo probabilístico con el objetivo de recolectar información de personas de más de cinco años de edad que residen los hogares ecuatorianos. En este sentido, la población objetivo son todas las personas de viviendas particulares de todas las provincias del Ecuador, exceptuando a Galápagos debido al nivel de representatividad INEC (2018b). Los individuos que se encuentran excluidos son la población establecida en viviendas colectivas, viviendas flotantes y sectores con población indigente.

El tratamiento de datos consistió en realizar un análisis exhaustivo de las variables obtenidas de la encuesta con el fin de hallar posibles inconsistencias e información faltante. El análisis evidenció que existe un 0.49% de datos perdidos con respecto a la pregunta sobre transacciones habituales realizadas en algún punto de atención financiero <sup>2</sup> y un 0.03% correspondiente a la pregunta sobre acceso a crédito <sup>3</sup>. De igual forma, las variables pobreza e ingreso per cápita tuvieron un 0.85% y 0.04% de datos perdidos, respectivamente. Con relación a la variable de edad, se eliminaron los individuos menores a 18 años y mayores a 85 años, los cuales corresponden al 1.90% de la muestra. Así mismo, con respecto a la variable de provincia, se eliminó el 0.51% de registros correspondientes a zonas no delimitadas. Por lo tanto, de la muestra original de 7201 registros se eliminaron 275, con lo cual la base usada en este estudio tuvo información sobre 6926 individuos. Por otra parte, varias variables fueron recategorizadas para su modelización y agrupadas en las cuatro categorías de acceso, uso, calidad y bienestar. La descripción detallada del proceso de depuración se encuentra en el anexo A.

Las variables usadas en la creación del indicador multidimensional fueron previamente clasificadas según la definición de inclusión financiera de Alianza para la Inclusión Financiera (2010) y sobre la cual se derivan los conceptos de acceso, uso, calidad y bienestar. De este modo, utilizando las preguntas de la sección financiera de la encuesta ENEMDU 2018 se crea-

---

<sup>2</sup>Corresponde a la pregunta fi04 de la encuesta ENEMDU DIC. 2018: ¿Qué transacciones realiza habitualmente en un punto de atención financiero?

<sup>3</sup>Corresponde a la pregunta fi14 de la encuesta ENEMDU DIC. 2018: Durante el último trimestre, ¿Usted solicitó un crédito?

ron cuatro categorías correspondientes a cada dimensión de inclusión financiera (Anexo C). La clasificación permitió que mediante la metodología propuesta se realice la inferencia de subindicadores sobre inclusión financiera. La tabla 3.1 muestra el conjunto total de variables y su correspondiente clasificación y descripción.

**Tabla 3.1:** Lista de Variables Usadas en la Creación del Indicador Multidimensional

Variable	Descripción	Tipo
Solicitó crédito	Si el individuo solicitó un crédito o no durante el último trimestre.	Acceso
No necesita productos	El individuo considera que no necesita un producto financiero.	Acceso
No confía instituciones	El individuo no gusta de productos financieros porque no confía en las entidades financieras.	Acceso
Cantidad pagos financieros	Cantidad de negocios cercanos al domicilio del individuo que aceptan formas de pago distintas al efectivo.	Acceso
No posee dinero	Si no posee productos porque no posee dinero para ahorrar financieramente.	Acceso
Dificultad apertura	El individuo considera que es difícil abrir una cuenta (trámites/documentos muy complejos).	Acceso
Cercanía punto atención	Cuán lejos está el punto de atención financiero más cercano para el individuo: Cerca, Lejos o Muy lejos.	Acceso

Continúa en la próxima página

**Tabla 3.1:** Lista de Variables Usadas en la Creación del Indicador Multidimensional(Continuación)

Solicitud aprobada	Si la persona solicito un crédito, este fue aprobado o no.	Acceso
Posee producto	Si posee o no alguna cuenta de ahorro, corriente, depósito a plazo o tarjeta de débito/crédito.	Acceso
Límites retiro	Individuo se siente insatisfecho porque sus productos tienen límites de retiro.	Bienestar
Costo mantener cuenta	Individuo se siente insatisfecho por los costos que implica mantener productos financieros.	Bienestar
Sucursal lejana	Individuo se siente insatisfecho porque la entidad financiera está lejos.	Bienestar
Posibilidad ahorrar	Individuo se siente satisfecho porque tiene posibilidades de ahorro con sus productos financieros.	Bienestar
Acceso otros productos	Individuo se siente satisfecho porque puede acceder a otros productos financieros.	Bienestar
Pagos seguros	Individuo se siente satisfecho por poder hacer pagos seguros.	Bienestar
Disponibilidad dinero.	Individuo se siente satisfecho por disponer de dinero de forma rápida.	Bienestar
Disponibilidad cajeros	Entidad tiene cajeros a disponibilidad del individuo.	Calidad

Continúa en la próxima página

**Tabla 3.1:** Lista de Variables Usadas en la Creación del Indicador Multidimensional(Continuación)

Califica punto atención		Cómo califica la persona a los productos financieros que usa.	Calidad
Disponibilidad correspondal no bancario	correspon- sal no bancario	Entidad tiene correspondal no bancario a disponibilidad del individuo.	Calidad
Disponibilidad agencia	sucursal	Entidad tiene sucursal de agencia a disponibilidad del individuo.	Calidad
Disponibilidad cooperativa	sucursal	Entidad tiene sucursal de cooperativa a disponibilidad del individuo.	Calidad
Frecuencia retiro		El comportamiento que la persona tiene para retirar dinero en una entidad o punto financiero.	Uso
Cantidad de servicios		Número de servicios financieros que la persona usa.	Uso
Recibe pago financiero		Si sus ingresos los recibe principalmente por medios financieros.	Uso
Frecuencia uso		Frecuencia con la que usa sus productos financieros en el mes.	Uso
Solicitud crédito formal		Si la persona solicitó o no un crédito a una entidad financiera.	Uso
Utiliza punto atención		Si utiliza o no algún punto de atención financiero para realizar alguna transacción.	Uso

Continúa en la próxima página

**Tabla 3.1:** Lista de Variables Usadas en la Creación del Indicador Multidimensional(Continuación)

*Fuente: ENEMDU Diciembre 2018*

*Elaboración: Los autores*

Por otro lado, para analizar los factores que influyen en los niveles de inclusión financiera se seleccionaron variables sociodemográficas, económicas y financieras con relación a la literatura. La tabla 3.2 muestra las variables seleccionadas, descripción y valores que pueden tomar.

**Tabla 3.2:** Lista de Variables Modelos Logit

Variable	Descripción
Área	Ubicación geográfica: Rural o Urbana.
Sexo	Hombre o Mujer.
Edad	Categoría de edad: <i>A: edad &lt; 26</i> (jóvenes) <i>B: <math>26 \leq edad \leq 59</math></i> (mediana edad) <i>C : edad &gt; 59</i> (adulto mayor)
Ingreso per cápita	Categoría de ingreso per cápita (USD): <i>A: Ingpc &lt; 85</i> (Pobreza) <i>B: <math>85 \leq Ingpc &lt; 163</math></i> <i>C : <math>163 \leq Ingpc &lt; 283</math></i> <i>D : Ingpc &gt; 283</i>
Condición de actividad	Pleno empleo, Subempleo, Desempleo y PEI.
Estado civil	Separado o Divorciado, Viudo, Unión libre, Soltero y Casado.

Continúa en la próxima página

**Tabla 3.2:** Lista de Variables Modelos Logit(Continuación)

---

Nivel de instrucción	Ninguno, Primaria, Secundaria y Superior.
Miembros del hogar	Cantidad de miembros en el hogar del jefe de hogar.
Etnia	Etnia con la cual se identifica el individuo según su cultura y costumbres.
Servicios básicos	La vivienda posee servicios básicos o no.
Vivienda propia	Si la vivienda es propia o no.
Seguro privado	Si el jefe de hogar cuenta con seguro privado o no.
Situación crediticia	Situación con respecto al último crédito que solicitó en una entidad financiera formal: Aprobado, Rechazado, No solicitó porque no puede pagar, No solicitó porque no necesita, Solicitó crédito informal.
Situación de ahorro	Situación con respecto al modo de ahorro de la persona: Ahorro formal, Ahorro informal y No ahorra.
Califica atención	Cómo califica la atención en los puntos de atención financieros: Bueno, Regular, Malo y No usa.
Situación bienestar	Está satisfecho con los servicios de su cuenta de ahorro, cuenta corriente, depósitos a plazo o tarjeta de crédito o débito: Satisfecho, No satisfecho y No posee.

---

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

## 3.2. Estadística Descriptiva

La muestra de información seleccionada para el análisis considera a los jefes de hogar que respondieron las preguntas del módulo financiero de la ENEMDU correspondiente a diciembre 2018. En este sentido, la cantidad de individuos que conforman la base de datos utilizada es 6926. Las preguntas involucradas en el módulo financiero están relacionadas con el acceso y uso de productos financieros, las cuales son analizadas a continuación con respecto a variables socio-demográficas.

La pregunta 1 del módulo financiero informa sobre el acceso que los individuos tienen hacia al menos un punto de atención financiera. La tabla 3.3 muestra cómo se distribuyen las respuestas de la pregunta en relación con el área de vivienda de la persona. En el área urbana, el 77.80 % de individuos utiliza al menos un punto de atención financiera, mientras que el 22.20 % no lo hace. Así mismo, en el área rural el 67.91 % de personas afirman utilizar al menos un producto financiero, mientras que el 32.09 % no lo hace.

**Tabla 3.3:** Acceso de Productos Financieros por Área

	Respuesta		Porcentaje	
	Sí	No	Sí	No
Urbana	3451	985	77,80 %	22,20 %
Rural	1691	799	67,91 %	32,09 %

Información obtenida de la pregunta: ¿Utiliza usted algún punto de atención financiero para realizar alguna transacción?

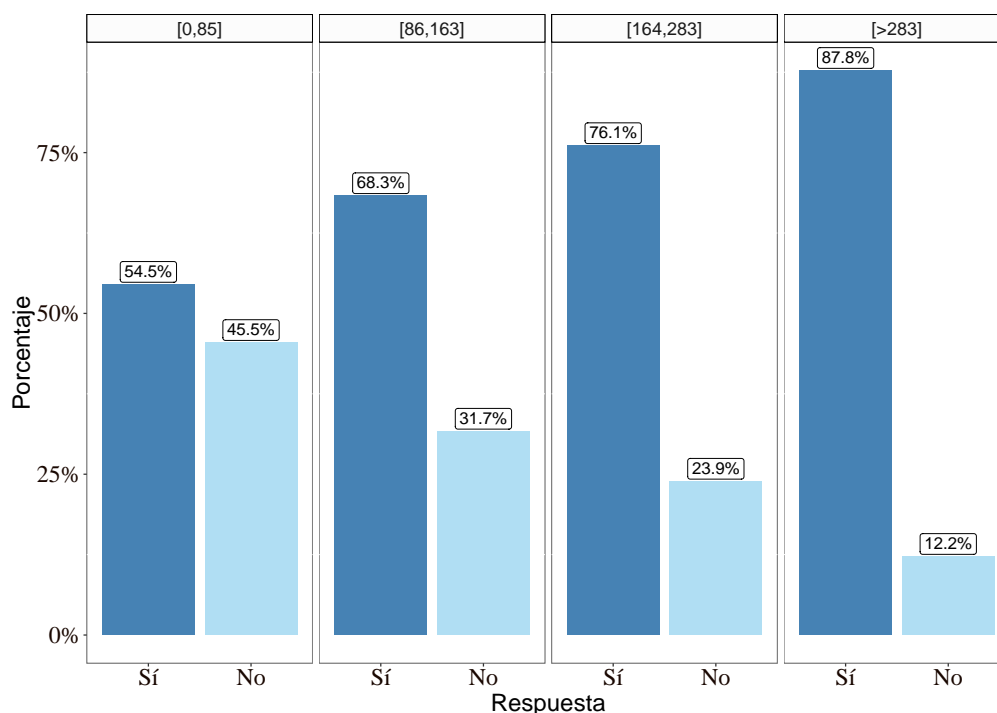
*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

Por otro lado, la figura 3.1 muestra cómo se relacionan las respuestas de la pregunta 1 con respecto al nivel de ingreso per cápita de las personas. Se observa que cada rango posee porcentajes de respuestas positivas mayores con respecto al rango anterior. De esta forma, el

porcentaje de individuos que afirma utilizar al menos un punto de atención financiero para cada rango de ingreso per cápita es de 54.5 %, 68.3 %, 76.1 % y 87.8 % respectivamente, en orden ascendente. Estos valores también implican que al aumentar el rango de ingresos existen cada vez menores porcentajes de individuos que no usan ningún punto de atención financiera.

**Figura 3.1:** Acceso de Productos Financieros por Rango de Ingreso Per Cápita



Información obtenida de la pregunta: ¿Utiliza usted algún punto de atención financiero para realizar alguna transacción?

*Fuente:* ENEMDU Diciembre 2018

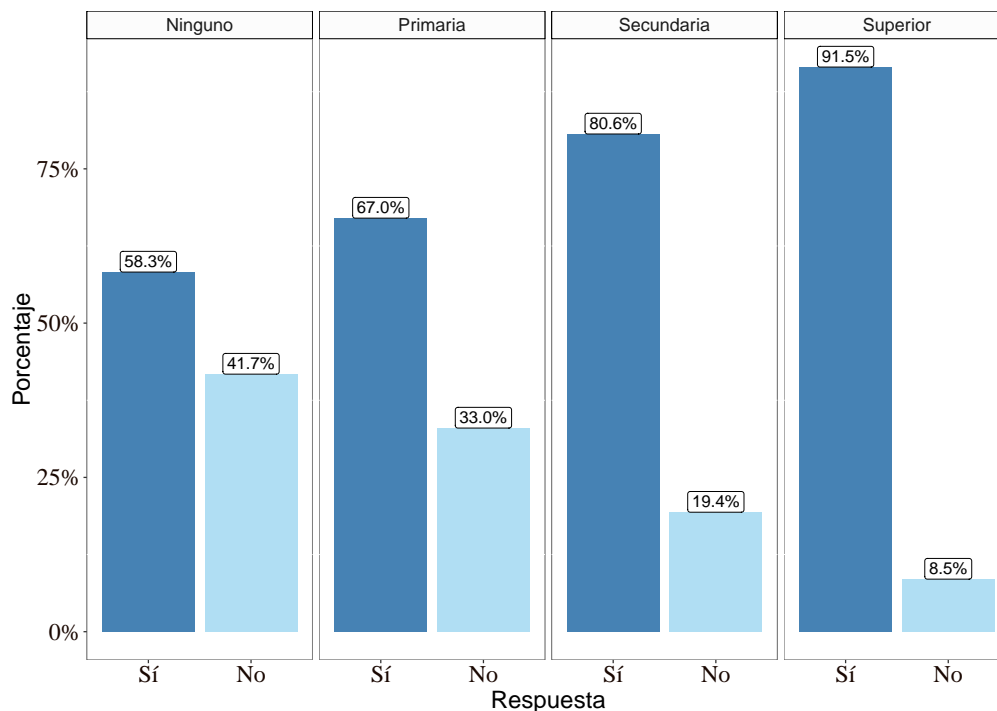
*Elaboración:* Los autores

De la misma forma, el gráfico 3.2 analiza la primera pregunta del cuestionario con respecto al nivel de instrucción. Se observa que existe un comportamiento similar al nivel de ingresos, es así como cada nivel de estudio posee un porcentaje de personas que usan al menos un punto de atención financiera mayor con el nivel anterior. En este sentido, el porcentaje de personas que no utilizan ningún punto de atención financiero es 41.7 %, 33 %, 19.4 % y 8.5 % respectivamente



para los niveles de estudio ninguno, primaria, secundaria y superior respectivamente.

**Figura 3.2:** Acceso de Productos Financieros por Nivel de Educación



Información obtenida de la pregunta: ¿Utiliza usted algún punto de atención financiero para realizar alguna transacción?

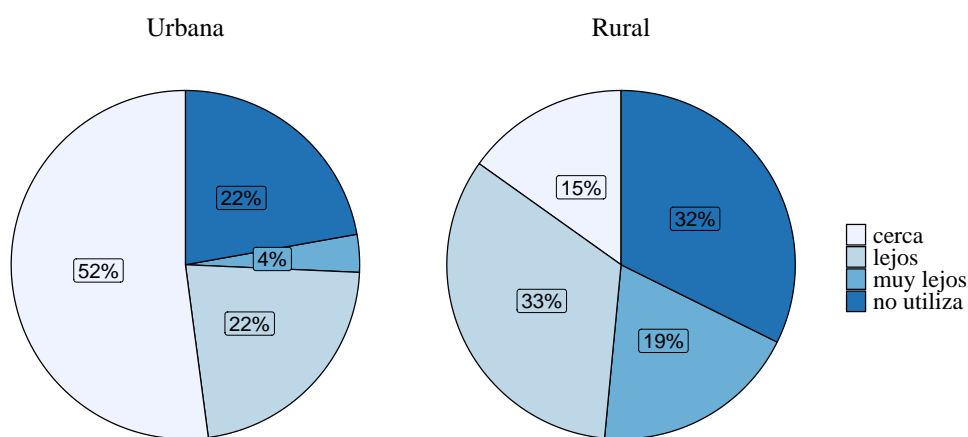
*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

Para realizar un análisis de acceso por cercanía al punto de atención financiero, se ha tomado la pregunta 2 del módulo de preguntas financieras y se ha dividido para el área de residencia del individuo. En primer lugar, en la figura 3.3 se verifica que el 22% de los individuos que viven en el área urbana indicaron no utilizar ningún punto de atención financiera. Mientras que en el área rural, el porcentaje aumenta en 10 puntos porcentuales, siendo un 32% de individuos los que mencionan no utilizar ningún punto de atención financiero. Asimismo, se observa que en el área rural, el mayor porcentaje indica encontrarse lejos y muy lejos al punto de atención financiera, siendo el 33% y 19% respectivamente. Mientras que solo el 15% indicó encontrarse

cerca. Por otra parte, en el caso de los individuos que residen en el área urbana, se observa que el 52 % indicaron encontrarse cerca al punto de atención financiera, mientras que el 22 % señaló encontrarse lejos y en un mínimo porcentaje del 4 % señalaron estar muy lejos.

**Figura 3.3:** Lejanía del Punto de Atención Financiero



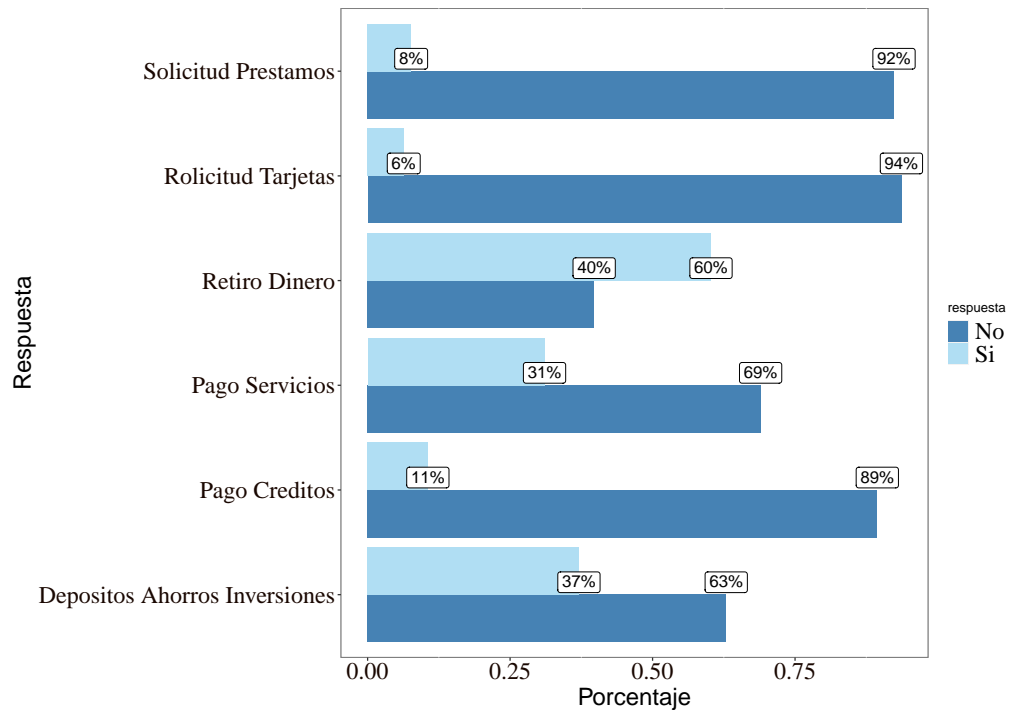
Información obtenida de la pregunta: ¿El punto de atención financiero que usted habitualmente utiliza para realizar transacciones se encuentra: Cerca, Lejos o Muy Lejos?

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

Para realizar un análisis del uso de productos o servicios financieros, en la figura 3.4 se puede observar que la mayoría de las transacciones que se realizan en Ecuador consisten en retiros de dinero, lo que representa el 60 % de los individuos. En segundo lugar, se encuentran los depósitos, ahorros o inversiones, que son realizados por el 37 % de los individuos, mientras que el 31 % realiza pagos de servicios.

**Figura 3.4:** Frecuencia de Uso en Puntos de Atención Financieros



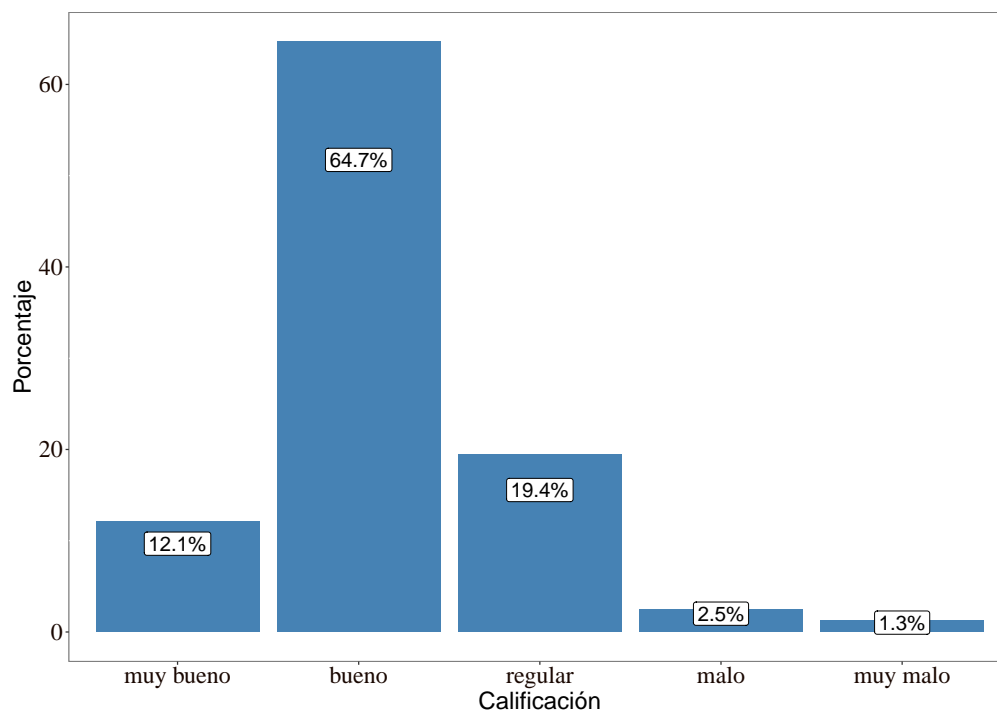
Información obtenida de la pregunta: ¿Qué transacciones realiza habitualmente en un punto de atención financiero?

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

En cuanto a la calidad de atención que reciben los usuarios en los puntos de servicios financieros, en la figura 3.5 se puede observar que el 12.1 % de los usuarios la califica como 'muy buena'. En un porcentaje mayor, el 64.7 % la califica como 'buena', seguido por el 'regular' con un 19.4 %. Finalmente, un porcentaje menor de usuarios, el 2.5 % y el 1.3 %, la calificaron como 'mala' y 'muy mala', respectivamente.

**Figura 3.5:** Calificación de Puntos de Atención Financieros



Información obtenida de la pregunta: ¿Cómo califica la atención en los puntos de atención financieros?

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

*Nota:* Se considera sólo a los individuos que sí utilizan los puntos de atención financieros

Por otro lado, la encuesta muestra información sobre individuos que no poseen ningún producto financiero. En este sentido, para dichas personas, la tabla 3.4 muestra el número de ellas que desearían contar con algún producto financiero. La cantidad de personas que respondieron de forma afirmativa fue 1042, mientras que la cantidad de personas que respondieron en forma opuesta fue de 1474. Esto representa el 41.42 % y 58.58 % respectivamente.

**Tabla 3.4:** Deseo de Productos Financieros

	<b>Respuesta</b>	
	<b>Sí</b>	<b>No</b>
Individuos	1042	1474
Porcentaje	41,42 %	58,58 %

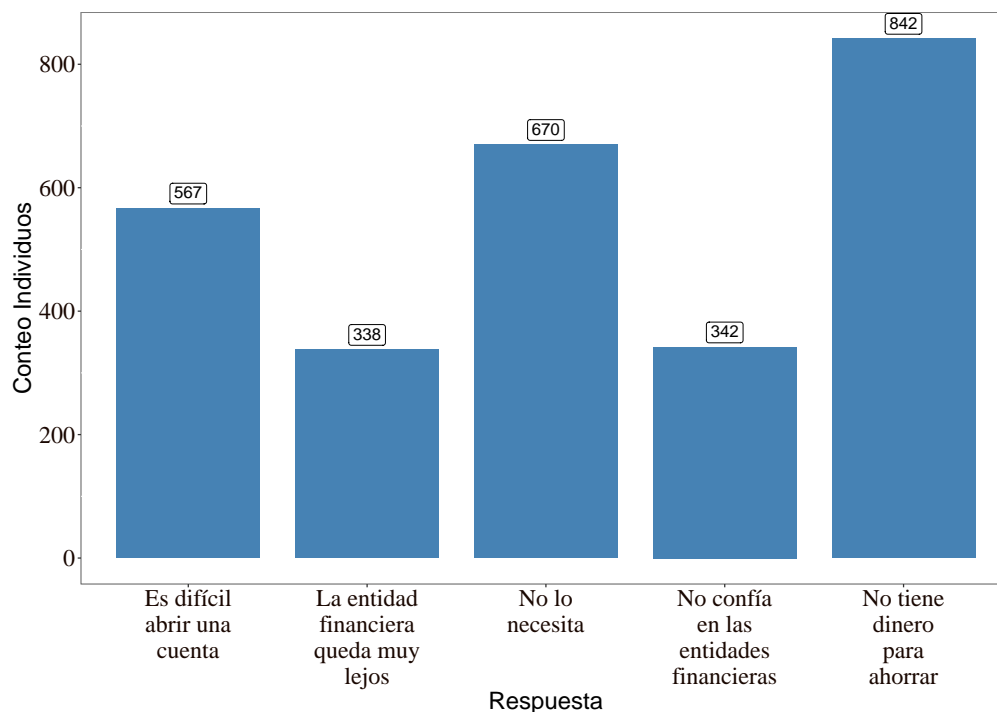
Información obtenida de la pregunta: ¿Le gustaría tener un producto financiero (cuenta ahorro/corriente, depósitos plazo, tarjeta crédito/débito)?

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

Así mismo, las razones por las cuales las personas no desearan tener ningún producto financiero se muestran en la figura 3.6. La razón que cuenta con un mayor número de respuestas fue ‘No tiene dinero para ahorrar’, seguida de ‘No lo necesita’, ‘Es difícil abrir una cuenta’ y ‘No confía en las entidades financieras’ respectivamente. Mientras tanto, que la de menor número de respuestas fue ‘La entidad financiera queda muy lejos’.

**Figura 3.6:** Desinterés Productos Financieros



Información obtenida de la pregunta: ¿Por qué no le gusta tener un producto financiero?

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

Por otro lado, con respecto a las personas que solicitaron un crédito durante el último trimestre, la tabla 3.5 muestra el número de respuestas para cada nivel de ingresos per cápita. Existen 256 personas en el rango de ingreso mayor a \$283 que solicitaron un crédito, seguido de 160 personas en el rango de \$164 a \$283, 131 personas en el rango de \$86 a \$163 y por último de 51 personas en el rango ingreso menor a \$85. Se observa que el número de solicitantes es mayor en cada rango de ingreso. El porcentaje de aprobación es de 86.27% para las personas que tienen un ingreso menor a \$85. El rango de ingreso de \$386 a \$163 tiene un porcentaje de aprobación, mientras que el de \$164 a \$283 uno del 90%. Por último, las personas del último rango tienen un porcentaje de aprobación de 90.63%.

**Tabla 3.5:** Aprobación de Crédito por Rango de Ingreso Per Cápita

Rango Ing. Per Cápita	Respuesta		Porcentaje	
	Sí	No	Sí	No
[0,85]	44	7	86,27 %	13,73 %
[86,163]	107	24	81,68 %	18,32 %
[164,283]	144	16	90,00 %	10,00 %
[>283]	232	24	90,63 %	9,38 %

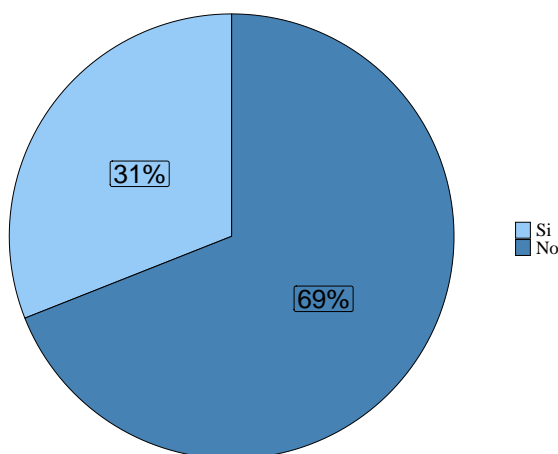
Información obtenida de la pregunta: ¿Del último crédito que solicitó en una entidad financiera lo aprobaron?

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

En cuanto al ahorro, en el gráfico 3.7 se puede visualizar que en el caso de Ecuador, el 69% de individuos no ahorra o guarda dinero, mientras que, solo el 31% sí lo hace.

**Figura 3.7:** Ahorro de Dinero



Información obtenida de la pregunta: ¿Usted ahorra o guarda dinero?

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

Al realizar un análisis por área, en la tabla 3.6 se puede observar que de aquellos individuos que residen en el área urbana, el 64.63 % señala no ahorrar o guardar dinero, mientras que, del grupo de individuos que reside en el área rural, el porcentaje es mayor siendo del 76.79 %.

**Tabla 3.6:** Ahorro de Dinero por Área

	Respuesta		Porcentaje	
	Sí	No	Sí	No
Urbana	1569	2867	35,37 %	64.63 %
Rural	578	1912	23.21 %	76.79 %

Información obtenida de la pregunta: ¿Usted ahorra o guarda dinero?

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores

Por otra parte, en la tabla 3.7 se pueden visualizar los resultados de ahorro por sexo, donde se verifica que existe una diferencia mínima entre hombres y mujeres. El 31.77 % de los hombres ahorran o guardan dinero, mientras que en el caso de las mujeres, el porcentaje es del 29.02 %.

**Tabla 3.7:** Ahorro de Dinero por Sexo

	Respuesta		Porcentaje	
	Sí	No	Sí	No
Hombre	1587	3409	31,77 %	68.23 %
Mujer	560	1370	29.02 %	70.98 %

Información obtenida de la pregunta: ¿Usted ahorra o guarda dinero?

*Fuente:* ENEMDU Diciembre 2018

*Elaboración:* Los autores



### **3.3. Metodología**

Amidžic et al. (2014) y Cámara y Tuesta (2014) utilizaron distintas metodologías para la creación de un indicador de inclusión financiera multidimensional, tales como el análisis factorial y el análisis de componentes principales. Por otro lado, en el presente este estudio se propone la metodología de *Variational Autoencoders (VAE)* para la construcción del indicador multidimensional y para la posterior evaluación de características sociodemográficas se usó un modelo *logit*.

Como se menciona anteriormente, la selección de variables financieras de la encuesta ENEM-DU de diciembre 2018 incluyó la categorización de cada una de estas según su relación con cada una de las cuatro dimensiones teóricas de inclusión financiera: acceso, uso, calidad y bienestar (tabla 3.1). Seguidamente, se agruparon a las variables en cuatro grupos para posteriormente ser utilizadas en la primera fase de la modelización VAE y crear los subindicadores correspondientes. Luego, cada uno de los subindicadores es agregado en la segunda fase mediante un nuevo modelo VAE para la creación del índice multidimensional.

Una vez construido el índice de inclusión financiera multidimensional se realiza un análisis para determinar su relación con las variables sociodemográficas y económicas seleccionadas (tabla 3.2). Para esto se crea la variable dependiente dicotómica con base en el indicador y se utiliza la metodología del modelo *logit* binario.

#### **3.3.1. Modelos de Variable Latente**

Muchos campos científicos suelen tratar con problemas de orígenes similares, uno de estos es la cuestión de representar datos de alta dimensionalidad en un espacio de menor tamaño. Estas representaciones pueden facilitar la visualización de patrones complejos y el descubrimiento de nuevas relaciones entre datos (Sorzano et al., 2014).

En las técnicas de reducción de dimensionalidad existen dos objetivos o componentes. El primero es la selección de características, el cual consiste en encontrar un subconjunto de variables perteneciente a otro conjunto más grande y que permita el modelamiento de un fenómeno usando un número menor de variables. El segundo es la extracción de características, el cual busca representar un conjunto de variables en otro de menor tamaño y que conserven las propiedades del primero (Maimon y Rokach, 2005). Para este segundo enfoque, algunos de los métodos más utilizados son el análisis de componentes principales (ACP) o el análisis discriminante lineal (ADL).

En este sentido, el problema de representar datos en una dimensión de menor tamaño es similar al de hallar un conjunto de variables no observadas que se relacionan directamente con el conjunto de variables observadas. Dichos factores no observados se denominan variables latentes y los modelos que las incorporan se denominan modelos de variables latentes (Cunningham y Yu, 2014). El uso de este tipo de modelamiento es útil para capturar relaciones complejas dentro de un sistema en el que son difíciles de medir directamente.

Los modelos de variables latentes son modelos probabilísticos que intentan explicar un cierto conjunto de datos ( $x$ ), generalmente de alta dimensión, en términos de otras variables no observadas ( $z$ ). Las variables no observadas o latentes son generadas mediante un proceso subyacente oculto. Este tipo de modelamiento es de gran utilidad en varias áreas y en especial en las ciencias sociales, en donde los problemas analizados a menudo contienen aspectos que no se pueden observar directamente. El origen de esta clase de modelos se da en el campo de la psicología con el trabajo de Spearman (1904) y su aplicación mediante análisis factorial. En la actualidad han cobrado una gran importancia por su capacidad para realizar inferencia sobre conceptos no observados, y mediante el desarrollo del aprendizaje automático<sup>4</sup>, la posibilidad

---

<sup>4</sup>El aprendizaje automático o *machine learning* es un término para referirse a un conjunto de algoritmos que aprenden de forma automatizada y son utilizados para resolver problemas o la toma de decisiones (Moreno et al., 1994).

de modelar estructuras complejas se ha robustecido (Lawrence, 2007).

En general, el modelamiento establece supuestos sobre la procedencia de las variables latentes y su relación con el conjunto de datos. De este modo, las variables latentes se modelan para seguir una distribución de probabilidad. En este sentido, en situaciones donde las variables no observadas ( $z$ ) son continuas, por lo general, se hace el supuesto de que se distribuyen normalmente. Así mismo, los parámetros del modelo indican la forma en que las variables latentes interactúan entre sí y también el modo en que las variables latentes se relacionan con las variables observadas (Lubke, 2018). La aplicación de este tipo de algoritmos permite la posibilidad de comprimir datos mientras se hacen supuestos mínimos o muy generales sobre su verdadera distribución (Van Der Maaten, 2009). Debido a que estos algoritmos infieren representaciones latentes directamente de los datos de entrada, son generalmente denominados como algoritmos de aprendizaje no supervisados.

En la inferencia de modelos latentes se suele tratar con formulaciones matemáticas intratables, es decir, que no se pueden resolver de forma directa (Berg y Kelly, 2006). Por ejemplo, al tratar de inferir la probabilidad  $P(x|z)$ , siendo  $z$  un conjunto de variables latentes, la ecuación  $P(x) = \int P(x,z)dz$  modela este fenómeno y se origina mediante la regla de Bayes. Esta integral se vuelve cada vez compleja por cada factor latente involucrado y, por lo tanto, imposible de evaluar (Tran et al., 2016). Es así que, ciertos problemas se vuelven intratables debido a su naturaleza y por ende se necesita una reformulación del mismo o uso de una metodología alterna.

En este sentido, diversas técnicas estadísticas permiten la inferencia aproximada de modelos intratables mediante la reformulación de uno nuevo, transformándolo así en un problema tratable. La formulación metodológica usada en este estudio está relacionada con un problema de difícil cálculo, por lo tanto, se usan los modelos de redes neuronales como herramienta auxiliar.

### 3.3.2. Redes Neuronales Artificiales

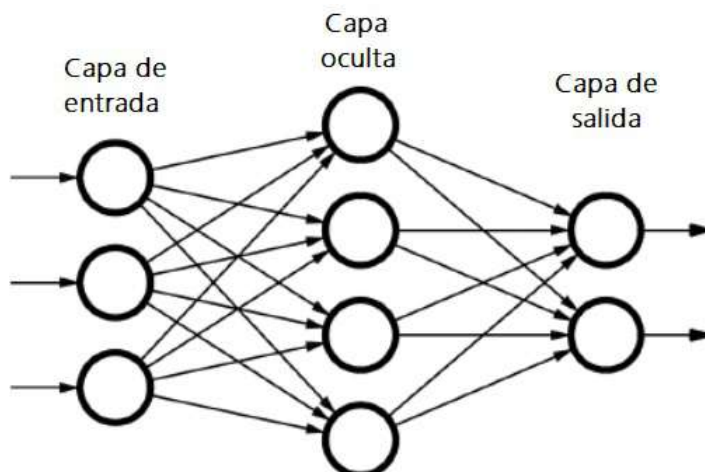
Las redes neuronales artificiales son un conjunto de algoritmos que buscan simular el funcionamiento del cerebro humano mediante un modelo matemático. Estos modelos están diseñados para buscar relaciones en conjuntos de datos mediante el uso de algoritmos de aprendizaje profundo<sup>5</sup>. Son comúnmente usadas para el reconocimiento de patrones, predicción o clasificación. Para este estudio, la utilidad en el uso de redes neuronales radica en su ayuda en la parametrización de una función auxiliar para la inferencia de una variable latente (Abiodun et al., 2019).

Una red neuronal artificial está conformada por elementos de procesamiento interconectados llamados neuronas, las cuales trabajan de manera conjunta para producir un valor de salida. Las redes neuronales están estructuradas en capas de entrada y salida, las primeras reciben los datos de entrada (input) mientras que las segundas entregan los datos procesados (output). También, poseen capas intermedias llamadas capas ocultas, las cuales transforman la entrada para poder ser procesada por la capa de salida (Scarselli et al., 2008). Cada capa está formada por un número determinado de neuronas. La figura 3.8 muestra la estructura básica que posee una red neuronal. La primera capa recibe los datos de entrada, mientras que la segunda los procesa y posteriormente son recibidos por la capa final, la cual procesa y entrega los datos de salida. En la práctica, una red neuronal puede poseer varias capas ocultas formadas por un distinto número de neuronas cada una.

---

<sup>5</sup>El aprendizaje profundo o *deep learning* es un enfoque del aprendizaje automático, el cual posee una estructura jerarquizada y en donde cada nivel tiene como función entender los datos que procesa de manera cognitiva (LeCun et al., 2015).

**Figura 3.8:** Estructura Básica de una Red Neuronal



*Fuente:* Buduma et al. (2022)

*Elaboración:* Los autores

En una red neuronal, los datos recibidos por cada neurona de una capa son procesados mediante dos funciones, una lineal y otra no lineal, con esto se obtienen resultados que son entregados a las neuronas de la siguiente capa para ser procesados nuevamente. De esta forma, la última capa de la red neuronal arroja los valores buscados para el problema a modelar. En tal sentido, este algoritmo busca modelar a las variables objetivo como una función no lineal de las variables de entrada y encontrar nuevas características (Goodfellow et al., 2016). Las variables de entrada contienen la información que reciben las neuronas de la primera capa en una red neuronal, y por ende la información inicial que es propagada a través de ella. Para el caso de este estudio, se plantea realizar cuatro modelos en los cuales la información inicial  $x_i$  de cada uno de ellos son las variables de cada uno de los cuatro conjuntos de datos correspondientes a acceso, uso, calidad y bienestar.

En este sentido, el proceso que sigue una neurona es definido por el siguiente algoritmo:

$$N = \phi \left[ \sum_i (x_i w_i) + b \right] \quad (3.1)$$

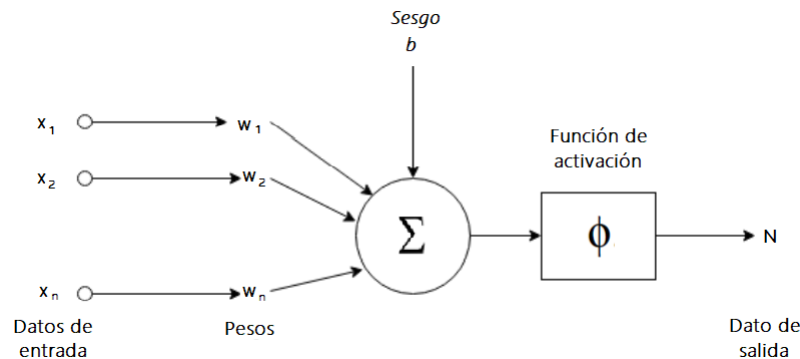
La ecuación 3.1 muestra el resultado  $N$  que obtiene cada neurona con base en la información  $x_i$  ingresada y su procedimiento se divide en dos partes. Primero, cada neurona procesa los datos entrada  $x_i$  multiplicándolos por un conjunto de parámetros  $w_i$  y luego sumándolos a un valor constante  $b$ . Este procedimiento refleja el modelamiento lineal de los datos de entrada. Los parámetros  $w_i$ , llamados también pesos, ajustan el valor de  $x_i$  mientras que el valor de  $b$  corrige el sesgo. Segundo, el resultado es introducido en una función no lineal predefinida  $\phi$  denominada función de activación<sup>6</sup>. Este paso permite que una red neuronal pueda modelar un sistema no lineal complejo mientras se combinan los datos de entrada de manera simple (Sharma et al., 2017). Es de esta forma que mediante la interconexión de las distintas neuronas de una red neuronal los valores de entrada son propagados hasta producir valores de salida.

El proceso de la ecuación 3.1 es representado en el gráfico 3.9. Se observa cómo los valores recibidos por una neurona son procesados, obteniendo así un valor de salida que es enviado a las neuronas de la siguiente capa. Este cálculo es realizado por cada neurona en la red, determinando así el rango numérico de las posibles salidas que se pueden tomar y controlando el flujo de información que recorre por la misma (Zaki et al., 2014). En general, el modelamiento de un problema mediante redes neuronales implica hallar el conjunto de pesos  $w$  que la conforman.

---

<sup>6</sup>La función de activación tiene como objetivo transformar el resultado de la combinación lineal de los pesos y los valores de entrada, estableciendo así el límite que tiene el valor de salida y poder proseguir a la siguiente neurona (Sharma et al., 2017).

**Figura 3.9:** Algoritmo Ejecutado por una Neurona



*Fuente:* Goodfellow et al. (2016)

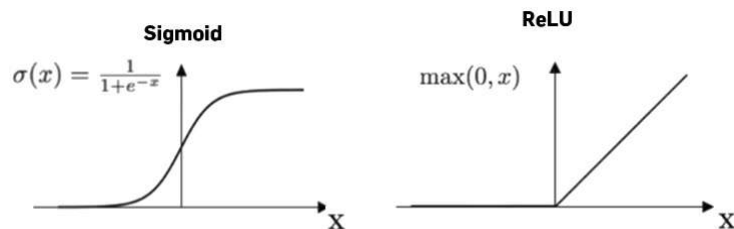
*Elaboración:* Los autores

Una red neuronal puede contener un número de parámetros a estimar muy alto dependiendo de la arquitectura que la defina. En este sentido, cada uno de los parámetros estimados influye en la salida que produce la red. Sin embargo, una diferencia importante entre un modelo de red neuronal y un modelo estadístico clásico radica en que los parámetros obtenidos por la red neuronal no son fácilmente interpretables. Dada la cantidad de pesos a estimar, no es posible observar de manera práctica como estos parámetros se relacionan con el conjunto de datos de entrada. Así, una red neuronal artificial es principalmente un modelo predictivo antes que un modelo explicativo (Ghorbani et al., 2019). En tal sentido, la transformación ejecutada por las neuronas hace posible que una red neuronal pueda resolver problemas complejos como clasificar imágenes o textos, generar nuevos datos o facilitar el descubrimiento de patrones (Zaki et al., 2014).

Entre algunas de las funciones de activación que pueden ser usadas tenemos a las funciones *ReLU*, *Sigmoide*, *TanH* y *Softmax*. Dependiendo del problema a resolver la elección de una u otra tendrá resultados más óptimos. Así mismo, cada capa de una red puede ser configurada para tomar distintas funciones de activación. En el caso de este estudio las funciones de activación

utilizadas son *ReLU* y *Sigmoide*, las cuales son representadas en la figura 3.10.

**Figura 3.10:** Funciones de Activación Utilizadas



*Fuente:* Goodfellow et al. (2016)

*Elaboración:* Los autores

Existen distintas clasificaciones para una red neuronal artificial, las cuales dependen de su número de capas, tipos de conexiones o el grado de las conexiones. El tipo de red más usada en cuanto a su arquitectura y la aplicada en este estudio es la red Neuronal *feed forward* o red neuronal prealimentada. Este tipo de red neuronal posee conexiones que van en una sola dirección, es decir, no forman un ciclo. Esta clase de redes son también denominadas perceptrón multicapa (MLP), ya que pueden tener varias capas ocultas formadas por varias neuronas (Aljalbout et al., 2018). La forma más simple de esta arquitectura de red neuronal es la presentada en la figura 3.8.

Por otra parte, una red neuronal artificial busca de forma implícita aproximar alguna función  $f^*$ . Así, por ejemplo, si el objetivo es clasificar un grupo de datos  $x$  teniendo una categoría de referencia  $y$ , su función de clasificación sería  $y = f^*(x)$ . En consecuencia, una red neuronal tiene la finalidad de crear una función  $y = f(x, \theta)$ , la cual es entrenada para aprender los parámetros  $\theta$ , de modo que se obtenga la mejor aproximación para la función objetivo (Ferrari, 2005). De este modo, el proceso de entrenar una red neuronal consiste en hallar los parámetros óptimos que conforman la red. En este sentido, la función objetivo que se busca modelar en este estudio es una función probabilística, la cual se detalla más adelante.



Una red neuronal artificial busca aprender a calcular la salida correcta para cada punto de datos mediante un algoritmo de aprendizaje. Esta fase se denomina proceso de entrenamiento y se realiza usando un conjunto de datos llamados de entrenamiento. Previo a este proceso se define el tipo de arquitectura que tendrá una red neuronal, la cual no cambia durante el entrenamiento, sin embargo, son los pesos de cada una de las conexiones las que pueden hacerlo. Por lo tanto, el aprendizaje de una red neuronal consiste en el ajuste o modificación de los pesos dada cierta información de entrada (Goodfellow et al., 2016). Normalmente, los pesos óptimos son hallados mediante la minimización o maximización de alguna función de costo o función de pérdida. Así, por ejemplo, una función de pérdida popular utilizada durante el entrenamiento es minimizar el error medio cuadrático, es decir, la diferencia al cuadrado del valor de salida y su valor esperado. (Scarselli et al., 2008). De igual forma, la elección de una función de pérdida depende del problema a tratar.

En la fase de actualización de los pesos de una red neuronal, por lo general no está involucrado todo el conjunto de datos, sino que se suele trabajar por lotes (batches) de datos. De esta forma, el conjunto total de datos es dividido en número  $D$  de lotes y se promedia el error generado, de modo que la actualización de pesos se enfoca en minimizar el error promedio. Este algoritmo permite que una red sea menos afectada por errores particulares en los datos y por ende mejora la calidad en los resultados (Goodfellow et al., 2016). Además, esto permite que el algoritmo sea más eficiente en cuanto a su cálculo. Cada lote de datos más pequeño se denomina mini-lotes (minibatch).

El uso de cada lote en el entrenamiento implica haber completado un ciclo y haber usado todo el conjunto de datos. En este sentido, cada recorrido por el conjunto de datos de entrenamiento completo se denomina *epoch* o época (Nielsen, 2015). Usualmente, para entrenar una red neuronal de manera óptima se necesitan varios ciclos o épocas. El propósito de este proceso es que la red neuronal tenga la capacidad de evaluar datos previos y con esto reajustar los

parámetros del modelo, de modo que no exista tendencia a ajustar el modelo con únicamente los últimos registros de datos. Es así que, en cada época, ocurren iteraciones en donde cada lote pasa por el algoritmo de aprendizaje (Nielsen, 2015).

Por otro lado, en la fase de entrenamiento el algoritmo que permite ajustar los pesos de manera iterativa se denomina propagación hacia atrás (backpropagation). El algoritmo *Backpropagation* ajusta los pesos de una red neuronal con base en los valores obtenidos por la función de pérdida obtenida en cada época anterior (Patterson y Gibson, 2017). Es así que, el error es propagado hacia atrás a través de la red neuronal desde la capa de salida. Con esto, los pesos de las neuronas de las capas ocultas cambian durante el entrenamiento. Esto hace que también los valores arrojados por la función de activación cambien y, por lo tanto, que cambie el valor de salida de una neurona.

El algoritmo *backpropagation* inicializa un valor aleatorio para los pesos  $w$  y lo propaga hacia adelante, es decir, recorre la red neuronal desde su capa de entrada. Después, dado que existe un error, se propaga hacia atrás y se corrige el valor de  $w$  con base en el lote de datos iterado. Los valores de los pesos  $w$  aumentan o disminuyen dependiendo de su influencia en la minimización del error. Por consiguiente, se busca obtener el valor de los pesos en el que el error sea mínimo. Si el error aumenta luego de una iteración, entonces su valor es propagado hacia atrás y se modifica el peso  $W$ .

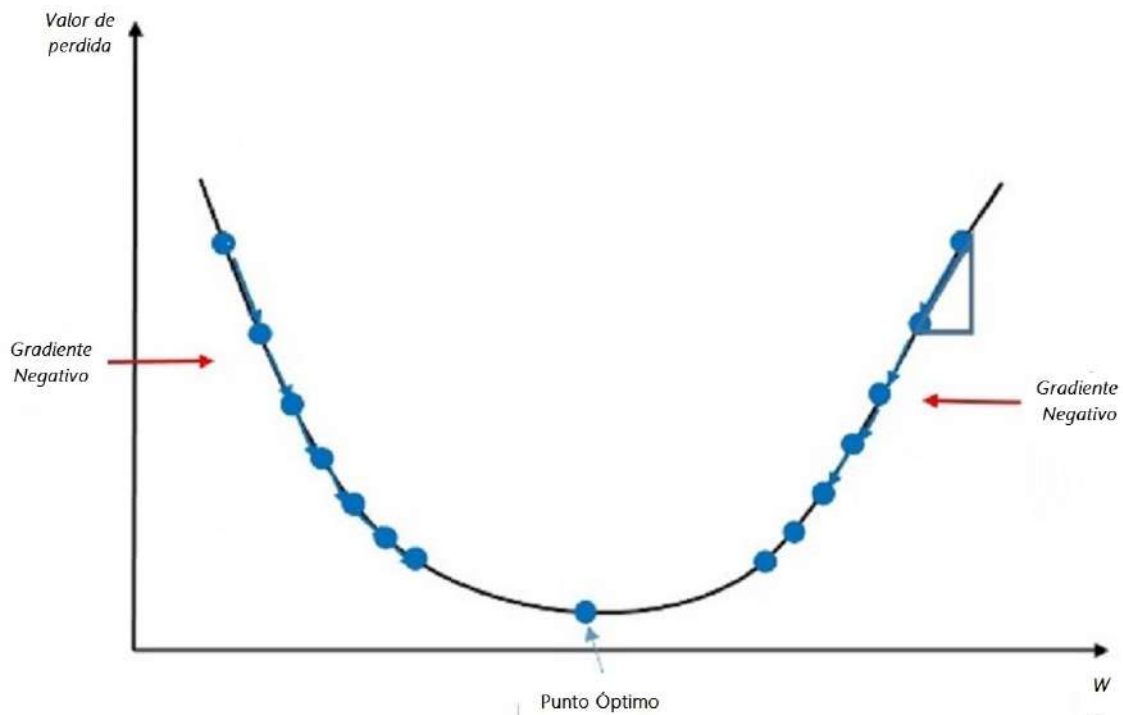
En la optimización de los pesos se encuentra involucrado un algoritmo de optimización llamado optimizador, el cual se encarga de obtener pesos mejores en cada iteración (Patterson y Gibson, 2017). El optimizador calcula el gradiente de la función de coste para cada peso de la red neuronal. Dado que se busca minimizar el error obtenido, los pesos se ajustan con base en la dirección negativa del gradiente tal que:

$$w_{t+1} = w_w - \eta \nabla_{\theta} f(\theta, x^{(i)}, y^{(i)}) \quad (3.2)$$

El término  $\eta$  de la ecuación 3.2 se denomina tasa de aprendizaje y controla como cambian los pesos en cada iteración. Así mismo, el par  $x(i), y(i)$  corresponde a un registro de datos del conjunto original o lote. De modo que un lote se forma tomando varios pares de datos. El tamaño del lote se define previo al entrenamiento y suele depender de la arquitectura de la red y datos que se dispongan (Goodfellow et al., 2016).

En general, el algoritmo *backpropagation* sigue un proceso como se detalla en la figura 3.11 tal que se alcance el mínimo valor de error.

**Figura 3.11:** Proceso del Algoritmo Backpropagation



Fuente: Goodfellow et al. (2016)

Elaboración: Los autores

El optimizador representado mediante la ecuación 3.2 se denomina *Stochastic Gradient Descent* (Descenso de Gradiente Estocástico) y es uno de los algoritmos más usados. En este sentido, existen varios optimizadores los cuales tienen como objetivo hallar el conjunto de pesos

que minimicen el valor de pérdida. Los optimizadores que se usaron en este estudio son: *SGD*, *Adam*, *Nadam*, *RMSprop* y *Adamax*.

### 3.3.3. Modelos Variational Autoencoders (VAE)

Un modelo *variational autoencoder* (VAE) es un algoritmo de aprendizaje semi-supervisado que combina redes neuronales mediante un enfoque probabilístico. Esta técnica pertenece al campo de los modelos generativos, los cuales son algoritmos de aprendizaje automático que buscan describir cómo un conjunto de datos es generado mediante un modelo probabilístico (Foster, 2019). El principal uso de la metodología VAE es el de construir modelos generativos que permitan la inferencia de datos sintetizados a partir de un gran conjunto de datos y que sigan sus mismos patrones.

Los modelos VAE son un tipo de red neuronal artificial que nacieron con base en el trabajo de Kingma y Welling (2013) y que sirvieron para formular una de las arquitecturas de aprendizaje profundo más conocidas para el modelamiento generativo. A menudo las redes VAE son usadas para la generación de nuevos datos con base en las características de los datos de entrada, y pueden ser alimentadas con cualquier tipo de dato: imágenes, textos, etc.

El modelo VAE se encuentra conformado por un codificador (*encoder*) y un decodificador (*decoder*). El codificador es una red neuronal que tiene como objetivo transformar los datos de entrada en una representación de menor dimensión y que posea sus propiedades principales, la versión reducida de los datos originales se dice es una representación latente de los datos de entrada. Por su parte, la representación latente obtenida por el codificador se utiliza como valor de entrada del decodificador, el cual tiene como objetivo transformarlo nuevamente a su estado original. Es así que, la arquitectura del *variational autoencoder* se forma por la unión del codificador y decodificador, los cuales son entrenados simultáneamente. Para este proceso el algoritmo usa una función de pérdida (función objetivo) que busca minimizar el error entre

la entrada del codificador y la salida del decodificador. Este proceso asegura que se penalicen datos de salida que sean distintos a los de entrada y por ende que la representación latente sea la más óptima posible (Klaas, 2019).

La modelización mediante un VAE resulta de principal interés por la utilidad que proporciona al crear una herramienta que permita la inferencia de una representación latente a través de su codificador (C. Zhang et al., 2018). Así mismo, la utilidad del decodificador radica en su potencial para la generación de nuevos datos, característica esencial de un modelo generativo. Ambos componentes son de igual importancia para el entrenamiento y su énfasis de estudio depende del problema a caracterizar (X. Zhang et al., 2019). En este trabajo el enfoque se centra principalmente en el codificador y sus capacidades para construir una representación latente.

El algoritmo *variational autoencoder* es construido basándose en la modificación del algoritmo denominado simplemente autoencoder y que tuvo sus primeros inicios en los 80 con la aplicación de redes neuronales y el trabajo de Ackley et al. (1985). La derivación de un modelo VAE es inferido como sigue: De acuerdo con Kingma et al. (2019), siendo  $x$  un set de datos de entrada (observados) con distribución de probabilidad  $P(x)$  desconocida y  $z$  una variable latente (no observada) con probabilidad  $P(z)$  desconocida, la función de probabilidad de los elementos de  $x$  se define como:

$$P(x) = \int P(x, z) dz \quad (3.3)$$

En donde  $z$  es una variable continua, mientras que  $x$  puede tomar valores continuos o discretos. De este modo, si, por ejemplo,  $z$  toma valores entre el intervalo  $[0, 1]$  y  $x$  toma valores reales en ese intervalo, entonces la distribución es continua (Doersch, 2016). Así mismo, si  $z$  es continua, la probabilidad  $P(x)$  será una mezcla infinita de distribuciones Gaussianas o también denominadas distribuciones de probabilidad compuesta (Rasmussen, 1999). De esta forma, la ecuación 3.3 parte de la probabilidad conjunta entre  $x$  y  $z$  para determinar la probabilidad de

ocurrencia del subconjunto  $x$  o probabilidad marginal.

De este modo, el modelamiento consiste en usar como valores de entrada  $x$  el conjunto de datos mostrados en la tabla 3.1, y con ello se busca obtener su representación latente correspondiente. En este sentido, siendo  $x$  cada uno de los conjuntos de variables que contienen las características de acceso, uso, calidad y bienestar para cada individuo en la muestra de datos; la ecuación 3.3 refleja la probabilidad de que un individuo tome determinados valores en  $x$  dado un valor determinado de  $z$ . Es así que, la variable oculta  $z$  es interpretada como el nivel en dicha dimensión que posee la persona en el sistema financiero y por ende que influye en su nivel general de inclusión financiera. De esta forma, los valores de cada variable oculta  $z$  están asociados a los valores que tome cada variable en el conjunto  $x$ .

Luego, la distribución conjunta de la ecuación 3.3 puede ser escrita como:

$$P(x, z) = P(z)P(x|z) \quad (3.4)$$

Así mismo, por las propiedades bayesianas básicas se tiene que:

$$P(z|x) = \frac{P(x|z)P(z)}{P(x)} = \frac{P(x|z)P(z)}{\int P(x, z)dz} \quad (3.5)$$

En donde  $P(z|x)$  es la distribución a posteriori;  $P(x|z)$  es la distribución marginal;  $P(z)$  es la distribución a priori y  $P(x)$  se dice es la evidencia.

El problema general de hallar la función  $P(x)$  responde a una estimación por máxima verosimilitud, la cual busca hallar un conjunto de parámetros para maximizar la probabilidad según los datos observados. Por lo tanto, se busca aproximar la distribución real de los datos mediante un modelo parametrizado  $P_{\theta}(x)$  con parámetros  $\theta$ . De modo que reescribiendo 3.3 usando 3.4 y 3.5 se tiene:

$$P(x) = \int P_{\theta}(x|z)P(z)dz = \int P_{\theta}(x, z)dz \quad (3.6)$$

Dado que la distribución de la variable latente  $z$  es continua, el cálculo de la integral en 3.8 es costoso computacionalmente. De igual forma, dado que  $P(x)$  no puede ser calculado (es intratable), tampoco lo es  $P(z|x)$  mediante el teorema de Bayes.

Para solucionar la intratabilidad de la distribución posterior se aproxima la misma mediante la técnica de inferencia variacional<sup>7</sup>. Esta técnica busca modelar directamente la distribución posterior mediante el uso de familias de distribuciones  $Q_{\Phi}$  con parámetros  $\Phi$ .

El objetivo es ajustar  $Q_{\Phi}$  con datos de muestra para aprender los parámetros  $\theta$ . Al seleccionar  $Q_{\Phi}$  se busca contar con una función fácil de manipular. Por ejemplo, su esperanza y constante de normalización pueden ser calculadas directamente a partir de los parámetros de distribución original. Por lo tanto, se puede usar  $Q_{\Phi}$  en lugar de  $P_{\theta}$  para hacer cualquier inferencia estadística (Blei et al., 2017).

Este modelo se denomina codificador o modelo de reconocimiento, mientras que los parámetros  $\Phi$  son denominados parámetros variacionales. Es así que buscamos optimizar los parámetros variacionales  $\Phi$  tal que:

$$Q_{\Phi}(z|x) \approx P_{\theta}(z|x)$$

Mediante el nuevo término y usando la ecuación 3.3 se puede aproximar la distribución  $P(z|x)$ :

$$\log P_{\theta}(x) = \log \int P(x, z)dz = \log \int P(x, z) \frac{Q(z)}{Q(z)} dz = \log(E_{Q_{\Phi}} \frac{P(x, z)}{Q(z)}) \quad (3.7)$$

En la práctica, para inferir los parámetros  $\theta$  y  $\Phi$  se usan redes neuronales artificiales, de

---

<sup>7</sup>La inferencia variacional es parte de los métodos estadísticos para aproximar integrales intratables y son usualmente aplicables a cualquier modelo de aprendizaje automático (Blei et al., 2017).

modo que se realiza una inferencia posterior sobre todos los puntos de datos. Este proceso se denomina inferencia variacional amortizada y a diferencia de métodos de inferencia variacional más tradicionales, los parámetros no se calculan por separado. Este proceso optimiza el proceso iterativo de inferencia y disminuye significativamente el número de parámetros a estimar (Agrawal y Domke, 2021).

La función de la ecuación 3.7 es conocida como función log-verosimilitud y es usada para inferir estimadores por el método de máxima verosimilitud. En el campo de la inferencia variacional no es este modelo logarítmico el que se evalúa por sí mismo, sino más bien una aproximación del límite inferior de él. En consecuencia, el objetivo de optimización del *variation autoencoder*, tal como en otros métodos variacionales, es el límite inferior de evidencia, a menudo abreviado ELBO o como límite inferior variacional (Blei et al., 2017).

De esta forma, mediante 3.7 obtenemos:

$$\log P_{\theta}(x) = E_Q[\log P_{\theta}(x)] \quad (3.8)$$

$$\log P_{\theta}(x) = E_{Q_{\phi}} \left[ \log \left[ \frac{P_{\theta}(x, z)}{P_{\theta}(z|x)} \right] \right] \quad (3.9)$$

$$\log P_{\theta}(x) = E_{Q_{\phi}} \left[ \log \left[ \frac{P_{\theta}(x, z) Q_{\phi}(z|x)}{Q_{\phi}(z|x) P_{\theta}(z|x)} \right] \right] \quad (3.10)$$

$$\log P_{\theta}(x) = E_{Q_{\phi}} \left[ \log \left[ \frac{P_{\theta}(x, z) Q_{\phi}(z|x)}{Q_{\phi}(z|x) P_{\theta}(z|x)} \right] \right] \quad (3.11)$$

$$\log P_{\theta}(x) = E_{Q_{\phi}} \left[ \log \left[ \frac{P_{\theta}(x, z)}{Q_{\phi}(z|x)} \right] \right] + E_{Q_{\phi}} \left[ \log \left[ \frac{Q_{\phi}(z|x)}{P_{\theta}(z|x)} \right] \right] \quad (3.12)$$



$$\log P_{\theta}(x) = \underbrace{L_{\theta, \phi}(x)}_{ELBO} + \underbrace{D_{KL}(Q_{\phi}(z|x) \parallel P_{\theta}(z|x))}_{Divergencia Kullback-Leibler} \quad (3.13)$$

El primer término en la ecuación 3.13 es límite inferior de evidencia (ELBO) y se observa como su valor detalla la diferencia entre el logaritmo de probabilidad real y la probabilidad aproximada. Por lo tanto, es un límite inferior de la probabilidad logarítmica  $\log P(X)$ . De modo que, el maximizar la probabilidad marginal se alcanza maximizando el límite inferior variacional (ELBO):

$$L_{\theta, \phi}(x) = E_{Q_{\phi}} [\log P_{\theta}(x, z) - \log Q_{\phi}(z|x)] \quad (3.14)$$

Así mismo, el segundo término en 3.13 es denominado divergencia *Kullback-Leibler (KL)* y es una medida de cómo una distribución de probabilidad difiere de otra distribución de probabilidad. Para este caso, KL mide la distancia desde la distribución aproximada  $Q_{\phi}$  a la verdadera distribución  $P_{\theta}$ . Este término es positivo y toma el valor de cero únicamente cuando la distribución  $Q_{\phi}$  es igual a la distribución posterior real

$$D_{KL}(Q_{\phi}(z|x) \parallel P_{\theta}(z|x)) \geq 0 \quad (3.15)$$

De la ecuación 3.13 y considerando la no negatividad de KL se concluye que  $L_{\theta, \phi}(x)$  es menor que la probabilidad logarítmica de  $x$ , es decir, es un límite (cota) inferior de la misma y por ello obtiene el nombre límite inferior de evidencia (ELBO).

$$L_{\theta, \phi}(x) = \log_{\theta} P(x) - \underbrace{D_{KL}(Q_{\phi}(z|x) \parallel P_{\theta}(z|x))}_{(+)} \quad (3.16)$$

$$\underbrace{D_{KL}(Q_\phi(z|x) \parallel P_\theta(z|x))}_{(+)} = \log_\theta P(x) - L_{\theta,\phi}(x) \quad (3.17)$$

$$L_{\theta,\phi}(x) \leq \log_\theta P(x) \quad (3.18)$$

La brecha entre el ELBO  $L_{\theta,\phi}(x)$  y la probabilidad marginal  $\log_\theta P(x)$  refleja que tan bueno es la aproximación de los datos. Mientras  $Q_\phi(z|x)$  se aproxime a la verdadera distribución (posterior)  $P_\theta(z|x)$  en términos de la divergencia KL, menor es la brecha.

En síntesis, la formulación de estos términos permite la inferencia de las funciones aproximadas y sus parámetros. Como herramienta auxiliar se usan las redes neuronales artificiales para la parametrización de estas funciones. El algoritmo de redes neuronales permite que se pueda optimizar de manera conjunta los parámetros  $\phi$  y  $\theta$  de la función objetivo, y, por lo tanto, la inferencia del espacio latente.

De esta manera, mediante 3.13 obtenemos la función objetivo a optimizar con base a cada lote de datos  $\mathcal{D}$ , en donde  $x$  es un punto de datos individual:

$$\mathcal{L}(\phi, \theta, \mathcal{D}) = \sum \mathcal{L}(\phi, \theta, x) \quad (3.19)$$

La función objetivo en 3.19 está conformada por la suma de los valores de pérdida individuales para cada registro de datos. De este modo, los parámetros  $\theta$  se encuentran mediante el gradiente  $\nabla_\theta$ :

$$\nabla_\theta \mathcal{L}(\phi, \theta, x) = \nabla_\theta E_{Q_\phi} [\log P_\theta(x, z) - \log Q_\phi(z|x)] \quad (3.20)$$

$$\nabla_\theta \mathcal{L}(\phi, \theta, x) = E_{Q_\phi} [\nabla_\theta (\log P_\theta(x, z) - \log Q_\phi(z|x))] \quad (3.21)$$

$$\nabla_{\theta} \mathcal{L}(\phi, \theta, x) \simeq \nabla_{\theta} [\log P_{\theta}(x, z) - \log Q_{\phi}(z|x)] = \nabla_{\theta} [\log P_{\theta}(x, z)] \quad (3.22)$$

Por otro lado, la derivación de los parámetros  $\phi$  mediante el gradiente se obtiene con  $\nabla_{\phi}$ :

$$\nabla_{\phi} \mathcal{L}(\phi, \theta, x) = \nabla_{\phi} E_{Q_{\phi}} [\log P_{\theta}(x, z) - \log Q_{\phi}(z|x)] \quad (3.23)$$

Sin embargo, dado que esta estimación se encuentra en función de los parámetros  $\phi$  desconocidos de  $Q_{\phi}(z|x)$ , se cumple que:

$$\nabla_{\phi} E_{Q_{\phi}} [\log P_{\theta}(x, z) - \log Q_{\phi}(z|x)] \neq E_{Q_{\phi}} [\nabla_{\phi} (\log P_{\theta}(x, z) - \log Q_{\phi}(z|x))] \quad (3.24)$$

Para solucionar este problema de estimación se introduce un método denominado "truco de reparametrización", el cual permite la estimación del gradiente mediante el descenso de gradiente estocástico en una red neuronal (Kingma et al., 2015). El método puede ser estimado dada la continuidad de la variable latente y la diferenciabilidad del codificador.

La técnica consiste en que la variable latente  $z$ , derivada del proceso  $Q_{\phi}(z|x)$ , se exprese mediante otra variable aleatoria  $\epsilon$ , la cual es diferenciable e invertible. Así mismo, esta variable no depende de  $x$  ni del parámetro  $\phi$ , tal que:

$$z = g(\epsilon, \phi, x) \quad (3.25)$$

Mediante la reparametrización se puede reemplazar la esperanza en función de  $Q_{\phi}$  en la ecuación del ELBO, derivada originalmente en 3.14, con una en función de la probabilidad de  $\epsilon$ .

De este modo, el algoritmo de descenso de gradiente estocástico dará valores iniciales a los

parámetros  $\theta$  y  $\phi$  y además lo hará con el nuevo parámetro  $\varepsilon$ . De modo que el ELBO se reescribe tal que:

$$\mathcal{L}_{\theta,\phi}(x) = E_{Q_\phi} [\log P_\theta(x, z) - \log Q_\phi(z|x)] \quad (3.26)$$

$$\mathcal{L}_{\theta,\phi}(x) = E_{P(\varepsilon)} [\log P_\theta(x, z) - \log Q_\phi(z|x)] \quad (3.27)$$

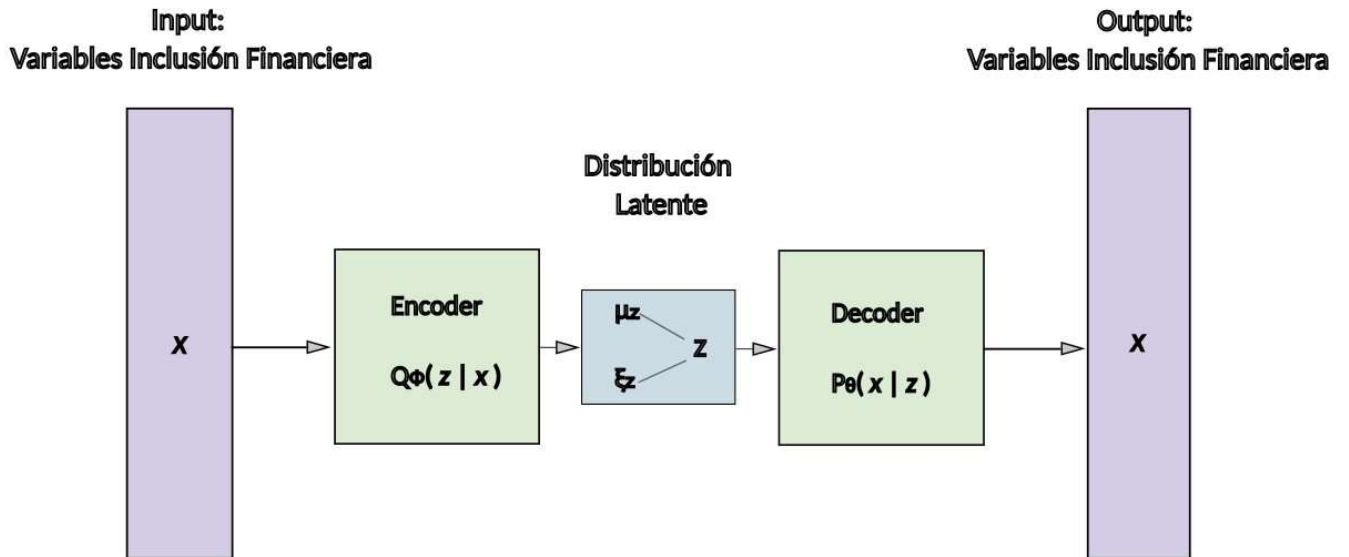
De esta forma, se halla la función objetivo aproximada  $\widehat{\mathcal{L}}_{\theta,\phi}(x)$  en dónde el término  $\varepsilon$ , denominado ruido, se infiere de su probabilidad  $P(\varepsilon)$ , tal que:

$$\varepsilon \sim P(\varepsilon) \quad (3.28)$$

$$\widehat{\mathcal{L}}_{\theta,\phi}(x)\varepsilon = \log P_\theta(x, z) - \log Q_\phi(z|x) \quad (3.29)$$

Finalmente, mediante la construcción de estos términos y el uso de redes neuronales artificiales para su parametrización, se construye el modelo que permite inferir la representación latente deseada. La figura 3.12 muestra la arquitectura general que tienen los modelos *variational autoencoder* planteados en este estudio. El input corresponde a cada conjunto de datos sobre inclusión financiera. Estos datos son procesados por cada elemento de la red. Así mismo, las conexiones representan la forma en que cada componente actúa. Los datos de entrada son procesados por el codificador generando la distribución del espacio latente, mismo con el cual se genera la variable latente correspondiente. Después, la variable latente ingresa en el decodificador para generar nuevamente los datos de entrada.

Figura 3.12: Estructura General Redes Neuronales VAE



Fuente: Kingma y Welling (2013)  
Elaboración: Los autores

La arquitectura de un modelo VAE tiene como objetivo que los datos de salida sean lo más similar posible a los datos de entrada. Con esto se busca que el modelo generativo sea capaz de reconstruir el input a partir de su representación latente. En este sentido; para cada conjunto de datos sobre acceso, uso, calidad y bienestar; se obtiene su representación latente correspondiente para posteriormente ser usada como subindicador de dicha dimensión.

El codificador tiene como salida dos vectores que contienen el conjunto de parámetros de la representación latente para cada punto de datos. Así mismo, se debe especificar el tamaño  $j$  que el espacio latente debe tener, para este estudio, en cada uno de los modelos  $j$  toma el valor de 1. Las variables latentes están restringidas a tener una distribución Gaussiana con su correspondiente matriz de covarianza. Por consiguiente, el codificador genera un vector que

contiene las medias,  $\mu$ , y el logaritmo de las varianzas,  $\log(\sigma^2)$  correspondiente a la variable latente de tamaño  $j$ . Por lo tanto, la salida del codificador toma la forma:

$$Q_{\phi}(z|x) = N(\mu_{\phi}(x), \log(\sigma^2)_{\phi}(x))$$

En donde el parámetro  $\phi$  es aprendido mediante redes neuronales.

De esta forma, habiendo aprendido los parámetros probabilísticos del espacio latente, se realiza un muestreo para cada uno de los puntos del conjunto de datos y de este modo obtener la representación latente  $z$ .

Por otra parte, la arquitectura del decodificador tiene como entrada la salida del codificador. De este modo, el objetivo del decodificador es reconstruir cada uno de los puntos de datos originales partiendo de su representación latente. Es así que, la distribución decodificadora es  $P_{\theta}(x|z)$ .

Con el fin de garantizar que la representación latente  $z$  sea una representación fiel de los datos de entrada y contenga información útil, el modelo VAE se entrena para minimizar una función de pérdida personalizada formada por dos componentes. Su primer componente es el error cuadrático medio de cada registro de datos para cada una de las variables que lo conforma. Cada término  $x_{i,j}$  es evaluado con su respectiva reconstrucción  $\hat{x}_{i,j}$  dentro de cada lote de tamaño  $M$ , mientras que el número de variables o factores es  $N$ .

Así el término de reconstrucción está dado por:

$$L_R(\phi, \theta, x_M, \varepsilon) = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^N (x_{i,j} - \hat{x}_{i,j})^2 \quad (3.30)$$

La reducción de 3.30 implica una buena reconstrucción, sin embargo, no garantiza que la distribución latente tome una forma Gaussiana. Para asegurar que la distribución de la variable latente se aproxime a la distribución a priori deseada, se introduce un segundo término en la

función de pérdida, el cual busca minimizar la divergencia Kullback-Leibler (KL). Dado el supuesto de una distribución Gaussiana a priori, el término de regularización de KL se calcula como:

$$L_{KL}(\phi, x_M) = \frac{1}{2M} \sum_{i=1}^M \sum_{j=1}^N (1 + \log(\sigma_{i,j}^2) - \mu_{i,j}^2 - \sigma_{i,j}^2) \quad (3.31)$$

Por consiguiente, la función objetivo a minimizar está dada por suma de los dos términos en 3.30 y 3.31.

$$L_{VAE}(\phi, x_M, \theta, \epsilon) = L_R + L_{KL} \quad (3.32)$$

De este modo, cada  $x_{i,j}$  corresponde a un registro de datos del conjunto correspondiente a cada una de las cuatro dimensiones de inclusión financiera.

En conclusión, para la obtención de cada subindicador  $SI_i$  se extrae el espacio latente de cada uno de los cuatro modelos VAE. Para este objetivo se hace uso del codificador y cada uno de los conjuntos de datos sobre inclusión financiera. De tal forma, cada subindicador se obtiene como:

$$SI_{acceso} = \text{CodificadorVAE}(x_{acceso})$$

$$SI_{bienestar} = \text{CodificadorVAE}(x_{bienestar})$$

$$SI_{calidad} = \text{CodificadorVAE}(x_{calidad})$$

$$SI_{uso} = \text{CodificadorVAE}(x_{uso})$$

Finalmente, mediante el uso de cada subindicador se agrega la información contenida en ellos con la ayuda de la metodología VAE para conformar el indicador de inclusión financiera multidimensional  $IF$ . Así mismo, se hace uso del codificador para la extracción del espacio latente.

$$IF = \text{CodificadorVAE}(SI)$$

En donde:

$$SI = [SI_{\text{acceso}}, SI_{\text{bienestar}}, SI_{\text{calidad}}, SI_{\text{uso}}]$$

### 3.3.4. Validación Modelos VAE

En la determinación de la consistencia de los modelos VAE para cada uno de los subindicadores e indicador general, se evaluó la estructura de las redes neuronales. La arquitectura de la red neuronal artificial describe la estructura general que posee la red y que será usada para la modelización. Esto implica especificar cuántas capas contiene la red y cuántas neuronas hay en cada capa. Así mismo, se debe especificar las funciones de activación y optimizadores utilizados.

Palmer et al. (2006) mencionan que, en general, una red neuronal artificial con una capa oculta y un número adecuado de neuronas ocultas tiene la capacidad de modelar bien cualquier función continua. Asimismo, se menciona que en la práctica las redes neuronales con un máximo de dos capas ocultas funcionan de forma adecuada para el objetivo deseado.

En la primera capa de entrada de cada red neuronal cada factor o variable está representado por una neurona. Por lo tanto, para el modelamiento de cada subindicador el número de neuronas en la capa de entrada será igual al número de variables independientes que se hayan definido para esa dimensión de inclusión financiera.



Teóricamente, la presencia de capas ocultas en una red neuronal aumenta la capacidad de la red para poder generalizar patrones, con lo cual se mejora la precisión de la red neuronal. Sin embargo, el poseer varias capas ocultas implica un mayor consumo computacional y, por ende, un mayor tiempo de procesamiento, además de tener una mayor posibilidad de sobreajuste del modelo (Shen et al., 2008)

En tal sentido, para la elección de la estructura de una red no existe una metodología estricta, sin embargo, a continuación, se resumen ciertos procesos heurísticos para su elección.

Bailey y Thompson (1990) recomiendan que el número de neuronas ocultas sea de  $0,75n$ , donde  $n$  es el número de neuronas de entrada. Así mismo, Katz et al. (1992) sugieren que el número de neuronas ocultas debe oscilar entre  $0,5n$  y  $3n$ , siendo  $n$  el número de neuronas de entrada. También, en el estudio realizado por Masters (1993) se propone que el número de neuronas vendría dado por  $0,5\sqrt{nm}$  y  $2\sqrt{nm}$ , siendo  $n$  el número de neuronas en la capa de entrada y  $m$  el número de neuronas de salida. Por otro lado, la investigación de Yoon y Swales (1991) sugiere que hay un número óptimo de neuronas ocultas a partir del cual el rendimiento de la red neuronal empieza a ser el mismo o incluso disminuir. Adicionalmente, Meireles et al. (2003) realizan un estudio basado en la prueba y error para determinar el número de neuronas. Su enfoque se basa en aumentar gradualmente el número de neuronas en varios modelamientos y con ello observar la configuración en la que el error disminuye.

Considerando estas premisas, para este estudio se han evaluado varias configuraciones con base en la estructura de la red y su elección de parámetros, con el objetivo de elegir la arquitectura que proporcione un mejor rendimiento. Para este fin, se define el método de búsqueda en rejilla (*grid search*), el cual consiste en definir un conjunto de posibles configuraciones en el que realizar la búsqueda (Liashchynskyi y Liashchynskyi, 2019). Para ello, se definen previamente los valores que puede tomar cada parámetro de la red neuronal y con se crean combinaciones de configuraciones. Luego se evalúa cada configuración y se determina un valor de la función

de pérdida final para cada una de ellas. Es así que, la configuración elegida es la que obtenga un menor valor de pérdida.

El posible número de neuronas se determina basándose en las especificaciones de Bailey y Thompson (1990); Katz et al. (1992); Masters (1993). Adicionalmente, se definieron configuraciones para el número de capas ocultas, el número de épocas, optimizador para la función de pérdida y tamaño del lote. Es así que, los posibles valores de parámetros definidos para cada modelo se definen como sigue:

- Modelo VAE Acceso

- I. Capa intermedia 1: 5,10,13
- II. Capa intermedia 2: 5,7,10
- III. Capa intermedia 2: 7,10
- IV. Épocas: 250,300
- V. Optimizador: RMSprop, Adam, SGD,Adamax
- VI. Tamaño Lote: 32,64

- Modelo VAE Bienestar

- I. Capa intermedia 1: 5,7
- II. Capa intermedia 2: 5,7
- III. Épocas: 250,300
- IV. Optimizador: RMSprop, Adam, SGD,Adamax
- V. Tamaño Lote: 32,64

- Modelo VAE Uso

- I. Capa intermedia 1: 5,12,15
  - II. Capa intermedia 2: 0,8,12
  - III. Épocas: 250,300
  - IV. Optimizador: RMSprop, Adam, SGD,Adamax
  - V. Tamaño Lote: 32,64
- Modelo VAE Calidad
    - I. Capa intermedia 1: 6,8
    - II. Capa intermedia 2: 6,8
    - III. Capa intermedia 3: 0,6,8
    - IV. Épocas: 250,300
    - V. Optimizador: RMSprop, Adam, SGD,Adamax
    - VI. Tamaño Lote: 32,64

Para cada uno de los modelos propuestos, el número de neuronas en la capa de salida y capa de entrada es el mismo que el número de variables de dicha dimensión evaluada. Para el caso de la parte codificadora, se cuenta siempre con una capa que contiene dos neuronas, las cuales funcionan como parámetros de media y desviación estándar correspondientes a la distribución de los datos de entrada. Esta capa se conecta finalmente con la capa que contiene la representación reducida del input (espacio latente). Por otra parte, el espacio latente que genera la parte codificadora funcionará como input para la parte decodificadora; por lo tanto, el número de neuronas de la capa de entrada para esta parte corresponde al tamaño del espacio latente, en todos los modelos este número es uno. Considerando esto, las estructuras restantes de cada red neuronal serán las posibles configuraciones antes mencionadas.

Por otro lado, según Kaastra y Boyd (1996) la función de activación sigmoidea es la más utilizada para datos financieros o de series temporales, ya que posee características valiosas como la no linealidad y la diferenciabilidad. Así mismo, Trippi y Turban (1992) menciona que la función sigmoidea es ideal para modelar situaciones basadas en el comportamiento. Por lo tanto, para los propósitos de esta investigación se seleccionó la función sigmoidea como la función de activación en cada capa de salida, la cual se define cómo:

$$\phi(z) = \frac{1}{1 + e^{-z}}$$

Habiendo seleccionado las configuraciones a evaluar, el siguiente paso es evaluar los valores obtenidos por la función de pérdida. Como se mencionó previamente, para el caso específico de las redes *variational autoencoder* la función de pérdida usada se basa en la divergencia *Kullback-Leibler* (Hall, 1987). Esta función permite la comparación entre dos distribuciones con el objetivo de verificar su similitud y en términos generales nos indica la esperanza de la diferencia de información entre ambas distribuciones. Es de tal forma que, los valores arrojados por la función de pérdida indicaran un mejor ajuste mientras más bajos sean, indicando así que los valores de *input* y *output* son altamente similares.

Por otra parte, para determinar qué factores específicos concernientes a cada dimensión contribuyeron más a la construcción del indicador obtenido por los VAE, se realizó un análisis mediante el método de valores Shapley. Esta técnica permite determinar el aporte promedio que tiene cada variable introducida en el modelo hacia la salida del mismo. Los valores de Shapley son un método que proviene de la teoría de juegos y que se basa en determinar cómo se distribuyen las ganancias y pérdidas de un juego a cada contribuyente del mismo. Su utilidad recae en que puede ser aplicado hacia cualquier modelo de aprendizaje automático y ayudar en la interpretación de los factores introducidos (Hart, 1989). Es así que, esta técnica permitirá evaluar los factores más importantes para cada subindicador y con ello se eliminarán los que

posean un bajo o nulo aporte.

### 3.3.5. Modelos Logit

En la siguiente fase de análisis, se describirá el modelo *logit* binomial utilizado en este estudio. Este modelo se ha utilizado ampliamente en la literatura para analizar y predecir eventos binarios. En el contexto de esta tesis, el modelo *logit* binomial se ha aplicado para analizar la probabilidad de que un individuo se encuentre incluido financieramente, en función de una serie de variables explicativas relacionadas con factores sociodemográficos, económicos y financieros. Para este objetivo se emplea el índice de inclusión financiera multidimensional creado en la primera fase mediante la metodología VAE.

En la creación de la variable dependiente se utilizó la técnica de análisis de grupos o *clusters* con el objetivo de crear grupos de individuos similares. Según Carrizosa et al. (2021), una variable continua debe ser categorizada mediante puntos de corte que permitan crear grupos lo más similares posible. En este sentido, el indicador multidimensional fue categorizado usando este análisis para crear dos grupos: incluido o no incluido financieramente, la forma de esta categorización se detalla en el anexo E.

Los modelos *logit* son un tipo de modelo estadístico utilizados para predecir la probabilidad de que ocurra un evento (Hilbe, 2009). El modelo *logit* se determina mediante la función logística y es una herramienta utilizada para modelar situaciones en las que hay dos resultados posibles. En tal sentido, la variable dependiente es de tipo dicotómico y se representa por la variable  $Y_i$ , la cual toma los valores de 1 o 0 para cada individuo de la muestra, según corresponda.

$$Y_i = \begin{cases} 1 & \text{si se encuentra incluido financieramente} \\ 0 & \text{si no se encuentra incluido financieramente} \end{cases}$$

En dónde  $Y_i$  es la variable explicativa que toma dos posibles valores con probabilidades  $\pi_i$  y

$1 - \pi_i$ .

Según Wooldridge (2015), este modelo puede ser expresado de la siguiente manera:

$$P(y = 1|x) = G(\beta_0 + \beta_1x_1 + \dots + \beta_kx_k) = G(\beta_0 + x\beta) \quad (3.33)$$

En donde  $G$  es una función que toma valores entre 0 y 1 para todos los números reales  $z$ , tal que:  $0 < G(z) < 1$ . Mientras que,  $x$  es un vector con variables explicativas y  $\beta$  un vector de coeficientes. Para el caso específico de un modelo logit, la función que asegura que las probabilidades se encuentren entre 0 y 1 es la función logística:

$$G(z) = \frac{\exp(z)}{1 + \exp(z)} \quad (3.34)$$

La función logística  $G$  es creciente y sus valores aumentan más rápidamente en  $z = 0$ . Así mismo,  $G \rightarrow 0$  cuando  $z \rightarrow -\infty$ , y  $G \rightarrow 1$  cuando  $z \rightarrow \infty$  (Wooldridge, 2015).

En tal sentido, para cada individuo de la muestra  $i$ ,  $z$  toma los valores:

$$z_i = \beta_0 + \beta_1x_{1i} + \beta_2x_{2i} + \beta_3x_{3i} + \dots + \beta_nx_{ni} + \varepsilon_i$$

Con:

$x_i$  = Conjunto de variables sociodemográficas, económicas, y financieras

$n$  = El número de variables

$\beta_0$  = Intercepto de la curva logística

$\beta_i$  = Pendiente de la curva logística

$\varepsilon_i$  = Término de error

Mediante esta especificación se estimarán cuatro modelos usando como variable dependiente el nivel de inclusión financiera y usando como variables independientes el conjunto de variables sociodemográficas y financieras detalladas en la tabla 3.2, tal que:

- $x_1$  : Área
- $x_2$  : Sexo
- $x_3$  : Grupo de edad
- $x_4$  : Ingreso per cápita (categórica)
- $x_5$  : Condición de actividad económica
- $x_6$  : Estado civil
- $x_7$  : Nivel de instrucción
- $x_8$  : Miembros del hogar
- $x_9$  : Etnia
- $x_{10}$  : Servicios básicos
- $x_{11}$  : Vivienda propia
- $x_{12}$  : Seguro privado
- $x_{13}$  : Situación crediticia
- $x_{14}$  : Situación ahorro
- $x_{15}$  : Califica atención
- $x_{16}$  : Satisfacción con los productos financieros

Los coeficientes  $\beta$  estimados permiten determinar si el efecto que tiene una variable independiente sobre de la variable dependiente es positivo o negativo. En tal sentido, para realizar la interpretación sobre la magnitud que tiene cada variable se deberá calcular los efectos marginales de las mismas (Wooldridge, 2015).

El efecto marginal que posee una variable independiente  $x$  se determina tal que:

$$\frac{\partial p(x)}{\partial x_i} = g(\beta_0 + \beta X)\beta_i \quad (3.35)$$

En dónde:

$$g(\beta_0 + \beta X) = g(z) = \frac{dG}{dz}(z)$$

La función  $G$  representa la densidad acumulada de una variable aleatoria continua, por lo tanto, la función  $g$  representa la densidad de probabilidad. En los modelos logit, la función  $G$  es estrictamente creciente, esto implica que la función de densidad de la probabilidad  $g(z)$  es mayor que cero para cualquier valor de  $z$ . Como resultado, el efecto parcial de la variable  $x_i$  sobre la probabilidad  $p(x)$  tiene el mismo signo que el coeficiente  $\beta_i$ , ya que depende de  $x$  a través de  $g(\beta_0 + \beta X)$  (Wooldridge, 2015).

Habiendo estimado cada uno de los modelos se procederá a realizar los correspondientes test de significancia estadística con el objetivo de observar la influencia que posee cada una de las variables y también establecer cuál de los modelos se ajusta de mejor forma para explicar el nivel de inclusión financiera.

### **3.3.6. Validación Modelos Logit**

Para la validación de cada modelo *logit*, se utilizaron diferentes medidas estadísticas para evaluar su capacidad de ajuste y discriminación, las cuales se encuentran en el Anexo D.

A continuación, se describen las medidas utilizadas y sus fundamentos teóricos.

#### **Correlación Policórica**

Para evaluar la presencia de multicolinealidad, se emplea la correlación policórica, una medida de asociación entre dos o más variables categóricas. A diferencia de otras medidas de asociación que asumen una relación lineal entre las categorías, esta medida considera la posibilidad de que la relación entre las categorías pueda ser no lineal. La correlación policórica se basa en una transformación de potencia de las categorías, que convierte las variables categóricas en valores numéricos continuos, y luego se utiliza una versión modificada del coeficiente de correlación de Pearson para calcular la correlación entre estas variables numéricas continuas (Hinkley, 1977).



$$r_p = \frac{1}{n-1} \sum_{i=1}^n \sum_{j=1}^n \left[ \left( \frac{U_{ij}}{n-1} \right)^{m-1} \right] \quad (3.36)$$

Donde:

$r_p$ : Coeficiente de correlación policórica.

$n$ : Número de observaciones en la muestra.

$m$ : Número de categorías para cada variable.

$U_{ij}$ : Número de coincidencias observadas entre las categorías  $i$  y  $j$  de las dos variables en estudio.

### Estadístico de la Devianza

El estadístico de la devianza  $D$  es definido como una función del logaritmo de la función de verosimilitud del modelo seleccionado y la del modelo saturado (modelo que se ajusta perfectamente a los datos) (Iglesias, 2013). Sigue una distribución  $\chi^2$  con grados de libertad igual a la diferencia entre la dimensión del espacio paramétrico y la dimensión del espacio bajo la hipótesis nula. De acuerdo con Collett (2002), la hipótesis nula que contrasta este estadístico es, que el modelo seleccionado estima perfectamente los datos observados y su expresión es la siguiente:

$$D = -2 \ln \left[ \frac{\widehat{\mathcal{L}}_c}{\widehat{\mathcal{L}}_F} \right] \quad (3.37)$$

Donde:

$\widehat{\mathcal{L}}_c$ : Valor de la función de verosimilitud estimada del modelo seleccionado.

$\widehat{\mathcal{L}}_F$ : Valor de la función de verosimilitud estimada del modelo saturado.

### **Estadístico de Hosmer Lemeshow**

La prueba de *Hosmer-Lemeshow* consiste en comparar los valores esperados por el modelo con los valores realmente observados. Ambas distribuciones, esperada y observada, se contrastan mediante una prueba  $\chi^2$ . La hipótesis nula de este test es que no hay diferencias entre los valores observados y los valores pronosticados (Hosmer Jr et al., 2013).

### **Curva ROC**

La curva ROC (*Receiver Operating Characteristic*) es una herramienta estadística que se utiliza para evaluar la capacidad discriminativa de una prueba diagnóstica dicotómica. Es una representación gráfica de la sensibilidad frente a la especificidad para distintos puntos de corte. Si se van modificando los puntos de corte y se representa la sensibilidad en el eje de las ordenadas frente al complementario de la especificidad en el eje de las abscisas, se obtiene la curva ROC J. M. Pérez y Martin (2023).

Considerando todos los puntos conjuntamente, se obtiene el área bajo la curva ROC (AUC), la cual toma valores entre 0 y 1, donde valores cercanos a 1 señalan un alto rendimiento del modelo (J. M. Pérez y Martin, 2023).

# Capítulo IV

## Resultados

En este capítulo se presentan los resultados sobre el modelamiento para la construcción de los subindicadores de acceso, bienestar, calidad y uso. Posteriormente, se detallan los resultados en la construcción del indicador multidimensional de inclusión financiera y se realiza un análisis cruzado con respecto a variables de las características de los individuos. Así mismo, utilizando el nivel de inclusión financiera como variable dependiente, se estiman modelos *logit* y se lleva a cabo un análisis e interpretación con respecto a variables sociodemográficas, económicas y financieras.

### 4.1. Resultados de los Modelos Variational Autoencoder

#### 4.1.1. Modelos VAE: Subindicadores

##### Arquitectura y valores de pérdida

En la presente investigación se estiman cinco modelos mediante la metodología VAE en dos etapas. Primero, los subindicadores de inclusión financiera (Acceso, Uso, Calidad y Bienestar) se construyeron mediante los cuatro primeros modelos. Después, el quinto modelo estima el

indicador multidimensional de inclusión financiera considerando los resultados obtenidos en la primera fase. En este sentido, se utilizó la técnica de rejilla (*grid search*) para determinar la arquitectura óptima de cada modelo con base en los hiperparámetros y variables seleccionadas (Liashchynskyi y Liashchynskyi, 2019). De esta forma, se obtuvieron los valores arrojados por la función de pérdida, los cuales sirven como criterio en la selección de modelos.

En relación con el modelo VAE de acceso, la tabla 4.1 muestra los primeros 10 resultados obtenidos con base al menor valor de pérdida alcanzado para cada configuración establecida. El máximo de estos valores es 0.14 856 mientras que el mínimo es 0,1345. En tal sentido, la arquitectura seleccionada para el modelo VAE de acceso es la presentada en la primera fila. El número de neuronas en la primera capa intermedia es 13, mientras que en la segunda es de 10 neuronas. Así mismo, para entrenar el modelo se usan 300 épocas, un tamaño de lote de 32 y el optimizador usado es RMSprop.

**Tabla 4.1:** Configuración VAE Acceso y Resultados de Pérdida

Épocas	Neuronas: Capa intermedia 1	Tamaño Lote	Neuronas: Capa intermedia 2	Optimizador	Valor Pérdida
300	13	32	10	RMSprop	0,13449942
300	5	64	3	Adam	0,1366709
300	15	64	5	Adam	0,13704279
300	13	64	5	RMSprop	0,13719502
300	13	64	5	Adam	0,13965219
300	15	32	5	Adam	0,14175239
300	5	32	3	Adam	0,14301005
300	13	32	7	Adam	0,14445145
300	15	64	5	RMSprop	0,14501951
300	15	32	7	Adam	0,14857806

*Elaboración:* Los autores

Por otro lado, la tabla 4.2 muestra los 10 mejores resultados de configuración para el modelo VAE de bienestar. Los valores de pérdida obtenidos estuvieron en el rango de 0,0992 a 0,1122. De igual forma, la configuración con menor pérdida se presenta en la primera fila y es la arquitectura seleccionada para esta dimensión. El número de neuronas para la primera y segunda capa intermedia es 5. Mientras que, para entrenar el modelo se usan 250 épocas, un tamaño de lote de 32 y el optimizador usado es Adam.

**Tabla 4.2:** Configuración VAE Bienestar y Resultados de Pérdida

Épocas	Neuronas: Capa intermedia 1	Neuronas: Capa intermedia 2	Tamaño Lote	Optimizador	Valor Pérdida
250	5	5	32	Adam	0,09915556
300	5	6	64	Adamax	0,10334451
300	5	9	32	Adamax	0,10528056
250	5	8	32	RMSprop	0,10839486
250	6	7	32	Adamax	0,10852379
250	6	4	32	RMSprop	0,10884284
300	6	9	32	RMSprop	0,11019092
300	6	6	32	Adamax	0,11204459
300	5	10	32	Adam	0,11209461
300	6	12	32	Adam	0,11224942

*Elaboración:* Los autores

Por su parte, los resultados para el modelo VAE de uso arrojaron valores de pérdida de entre 0,1322 a 0,1377. La tabla 4.3 muestra las primeras 10 mejores configuraciones y la elegida es la mostrada en la primera fila. El número de neuronas para la primera y segunda capa intermedia es 15. El número de épocas usado es de 250 con un tamaño de lote de 64 y el optimizador usado es Adam.

**Tabla 4.3:** Configuración VAE Uso y Resultados de Pérdida

Épocas	Neuronas: Capa intermedia 1	Neuronas: Capa intermedia 2	Tamaño Lote	Optimizador	Valor Pérdida
300	15	15	64	Adam	0,13218469
300	15	12	64	Adam	0,1328415
300	18	15	64	RMSprop	0,13351255
300	18	12	64	Adam	0,13400174
300	15	15	64	RMSprop	0,1343514
300	18	8	64	RMSprop	0,13466198
300	15	8	64	Adam	0,13547276
300	15	12	64	RMSprop	0,13650246
300	12	12	64	RMSprop	0,13651069
300	18	12	64	RMSprop	0,13769316

*Elaboración:* Los autores

Finalmente, los resultados para el modelo VAE de calidad tuvieron valores de pérdida en el rango de 0,1438 a 0,1524. Las 10 primeras mejores configuraciones se presentan en la tabla 4.4. El número de neuronas para la primera y segunda capa intermedia es de 8 y se cuenta con una capa adicional de 6 neuronas. El número de épocas usado es de 300 con un tamaño de lote de 64 y el optimizador usado es Adam.

**Tabla 4.4:** Configuración VAE Calidad y Resultados de Pérdida

Épocas	Neuronas: Capa intermedia 1	Neuronas: Capa intermedia 2	Tamaño Lote	Neuronas: Capa intermedia 3	Optimizador	Valor Pérdida
300	8	8	64	6	RMSprop	0,14376862
300	8	6	64	8	Adam	0,14609365
300	6	8	32	8	RMSprop	0,14658297
300	6	6	64	8	Adam	0,14680201
300	8	8	32	6	Adam	0,14876367
300	8	6	32	8	RMSprop	0,14971748
300	8	8	64	8	Adam	0,1519195
300	6	6	64	6	RMSprop	0,15199121
300	8	6	32	6	Adam	0,15208621
300	6	8	32	6	Adam	0,15243316

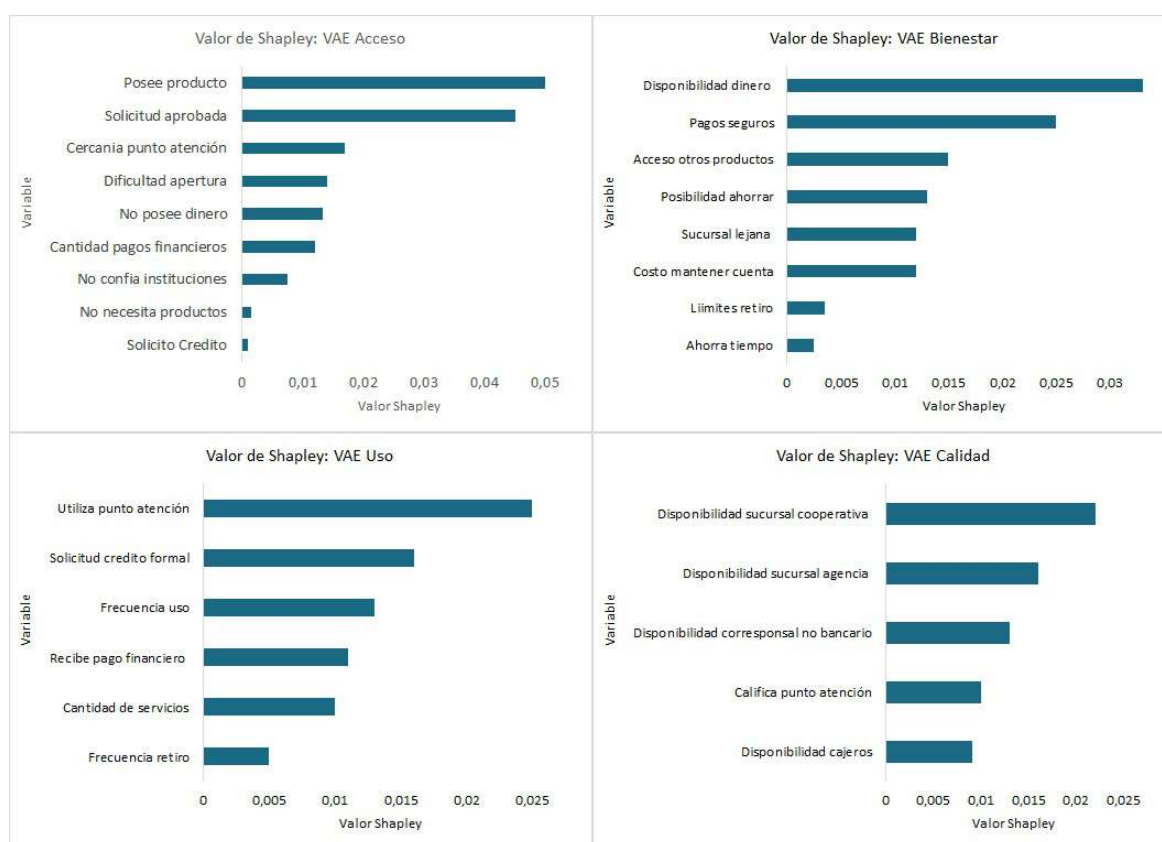
*Elaboración:* Los autores

A su vez, se evaluó la calidad en el aporte que cada variable usada en los modelos brinda a la salida de los mismos y por ende a la construcción de los espacios latentes. Para este objetivo se hace uso de los valores de Shapley introducidos en el capítulo 3, los cuales miden la contribución marginal promedio que tiene cada variable en los valores de salida de cada modelo. La figura 4.1 muestra para cada modelo VAE el valor Shapley obtenido por las variables que lo conforman.

En el modelo VAE de acceso la variable con mayor aporte es 'Posee producto' y 'Solicitud aprobada', así mismo las variables con menos aporte son 'No necesita productos' y 'Solicito Crédito'. Dado que el aporte de estas últimas variables es mínimo en relación con las otras variables, se eliminan del modelo. Por otra parte, en el modelo VAE de bienestar las variables con mayor aporte son 'Disponibilidad dinero' y 'Pagos seguros', mientras que las de menor aporte son 'Limites retiro' y 'Ahorra tiempo', las cuales son eliminadas dado su bajo aporte. Así

mismo, en el modelo VAE de uso las variables de mayor aporte fueron 'Utiliza punto atención' y 'Solicitud crédito formal', todas tienen un aporte significativo al modelo de forma que no se elimina ninguna. Finalmente, en el modelo VAE de calidad las variables de mayor aporte fueron 'Disponibilidad sucursal agencia' y 'Disponibilidad sucursal cooperativa', de igual forma todas las variables brindan un aporte significativo y no es eliminada ninguna.

**Figura 4.1:** Valor de Shapley: Modelos VAE



*Elaboración:* Los autores

De esta forma, las variables que conforman cada dimensión de inclusión financiera y que se utilizaron en el entrenamiento de modelos fueron:

- Acceso: No confía instituciones, Cantidad pagos financieros, No posee dinero, Dificultad



apertura, Cercanía punto atención, Solicitud aprobada, Posee producto.

- Bienestar: Costo mantener cuenta, Sucursal lejana, Posibilidad ahorrar, Acceso otros productos, Pagos seguros, Disponibilidad dinero.
- Uso: Cantidad de servicios, Recibe pago financiero, Frecuencia uso, Solicitud crédito formal, Utiliza punto atención.
- Calidad: Disponibilidad cajeros, Califica punto atención, Disponibilidad corresponsal no bancario, Disponibilidad sucursal agencia, Disponibilidad sucursal cooperativa.

### **Consistencia de modelos**

Por último, una vez seleccionada la arquitectura óptima de cada red y las variables que las conforman, se evalúa la consistencia de los modelos en sus salidas. La metodología de redes neuronales posee un componente probabilístico, por ende, se efectuaron varias ejecuciones o simulaciones en los modelos con el objetivo de analizar la estabilidad en el valor de pérdida resultante para los datos de entrenamiento y validación. En este contexto, cada conjunto de datos es separado en datos de entrenamiento y validación, el primer conjunto es usado para determinar los pesos óptimos de cada capa oculta y el segundo se usa para validar los resultados obtenidos por la red neuronal. De esta forma, la tabla 4.5 muestra quince ejecuciones distintas para cada dimensión de inclusión financiera y sus correspondientes valores obtenidos por la función de pérdida. Así mismo, cada una de las simulaciones muestra el valor promedio y varianza de los valores arrojados por la función de pérdida. El promedio de pérdida en cada modelo es muy cercano entre datos de entrenamiento y validación. También, la varianza en cada uno de ellos es de aproximadamente cero, mostrando así poca variación en los resultados obtenidos. De esta forma, se garantiza que cada uno de los modelos es consistente en sus representaciones latentes obtenidas.

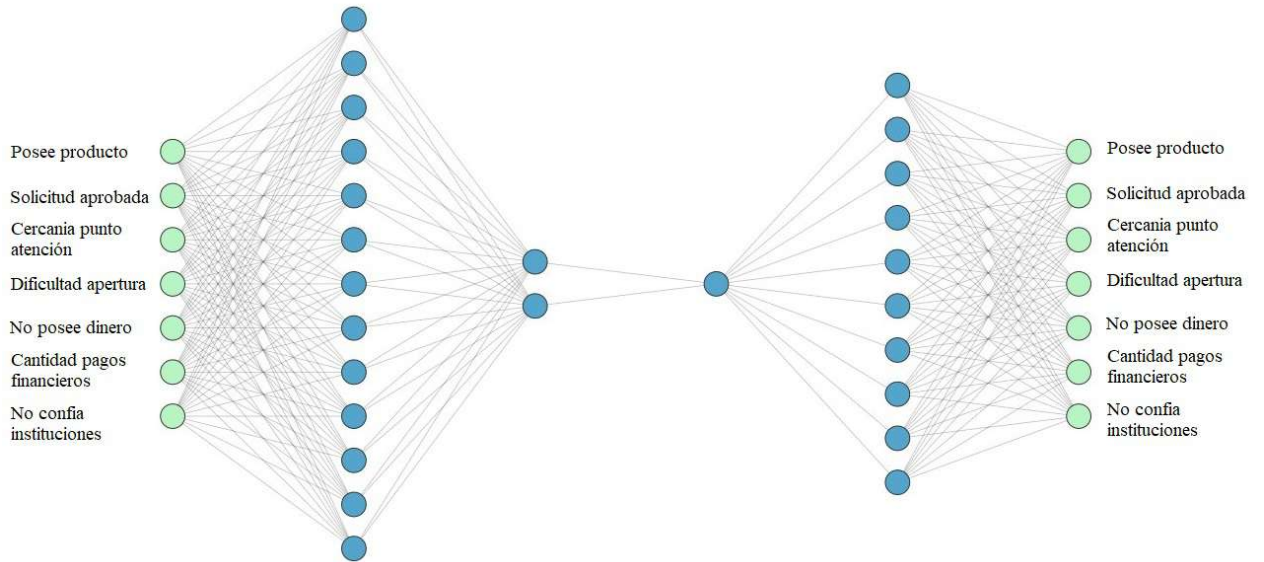
**Tabla 4.5:** Simulaciones Modelos VAE: Subindicadores

Simulación	Acceso		Bienestar		Uso		Calidad	
	Ent.	Val.	Ent.	Val.	Ent.	Val.	Ent.	Val.
1	0,12470	0,1388	0,12380	0,1235	0,1172	0,1186	0,1651	0,1608
2	0,12280	0,1199	0,14150	0,1491	0,1173	0,1171	0,1768	0,175
3	0,12670	0,1239	0,12340	0,1201	0,1189	0,1148	0,1544	0,1524
4	0,12000	0,1235	0,14720	0,1459	0,1295	0,1286	0,1565	0,1568
5	0,12500	0,1249	0,13600	0,1345	0,1647	0,1735	0,1531	0,1545
6	0,13290	0,1353	0,12940	0,1323	0,1272	0,1286	0,1542	0,1541
7	0,11240	0,1154	0,10580	0,1074	0,1233	0,1195	0,1557	0,1598
8	0,12120	0,1236	0,13200	0,1327	0,1233	0,1223	0,1571	0,1536
9	0,13800	0,1411	0,13140	0,1304	0,111	0,1148	0,1571	0,1537
10	0,12090	0,1163	0,15860	0,1593	0,1182	0,1142	0,153	0,1481
11	0,11270	0,114	0,12300	0,1243	0,1327	0,1421	0,1534	0,1519
12	0,11480	0,1167	0,11800	0,1179	0,1885	0,1885	0,1505	0,1501
13	0,13390	0,1559	0,16890	0,1698	0,132	0,1246	0,1553	0,2322
14	0,11630	0,1203	0,12800	0,1235	0,1178	0,1197	0,1539	0,1531
15	0,10630	0,1103	0,10120	0,1263	0,1094	0,1231	0,1531	0,1456
Promedio	0,12191	0,12533	0,13121	0,13313	0,12873	0,13000	0,15661	0,16011
Varianza	0,00006	0,00015	0,000320	0,000274	0,00045	0,00049	0,00004	0,00044

*Elaboración:* Los autores

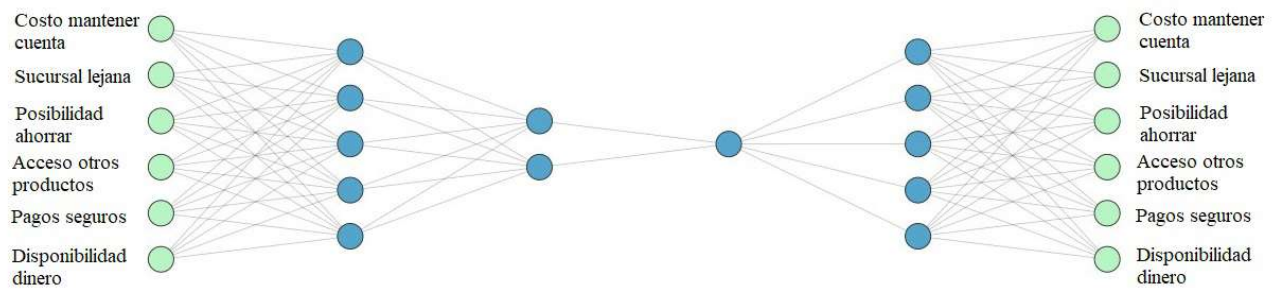
Finalmente, las figuras 4.2,4.3,4.4,4.5 reflejan de forma gráfica la arquitectura resultante que se usó para la inferencia de cada espacio latente correspondiente a cada dimensión de inclusión financiera. Además, el anexo B indica el valor de pérdida para cada época de entrenamiento. De esta forma, se verifica que en cada iteración final se alcanzan los pesos que permiten minimizar la función de pérdida, y asimismo se refleja cómo la pérdida obtenida para cada partición de datos tiende a ser de aproximadamente cero.

**Figura 4.2:** Estructura de Red Neuronal Modelo VAE Acceso



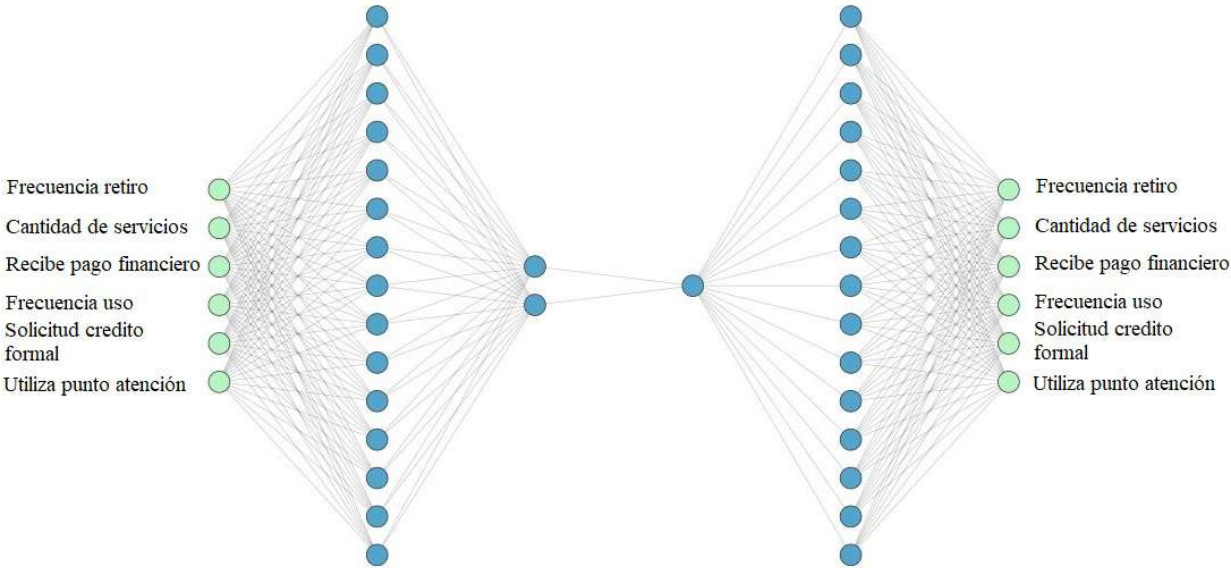
*Elaboración:* Los autores

**Figura 4.3:** Estructura de Red Neuronal Modelo VAE Bienestar



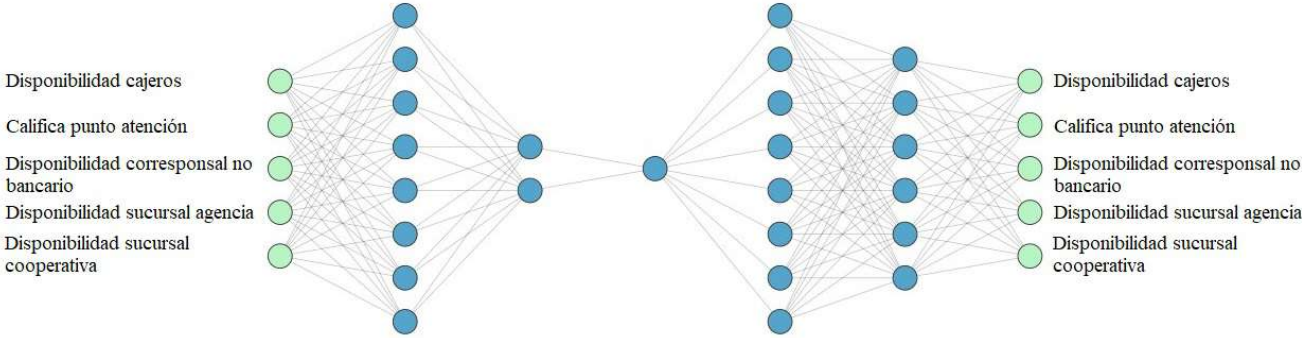
*Elaboración:* Los autores

**Figura 4.4:** Estructura de Red Neuronal Modelo VAE Uso



*Elaboración:* Los autores

**Figura 4.5:** Estructura de Red Neuronal Modelo VAE Calidad



*Elaboración:* Los autores

## 4.1.2. Modelo VAE: Indicador Multidimensional de Inclusión Financiera

### Arquitectura y valores de pérdida

Una vez que se ha entrenado cada uno de los cuatro primeros modelos VAE, se procede a obtener la representación latente de cada dimensión de inclusión financiera. Luego, para la construcción del indicador multidimensional de inclusión financiera, se usa cada una de las representaciones como variables de entrada durante el nuevo modelamiento VAE. De esta forma, el modelo final contiene cuatro valores de entrada y salida. Por su parte, para la evaluación de la configuración óptima de este modelo se utiliza la técnica de rejilla.

Con los hiperparámetros definidos para evaluar el modelo VAE, la tabla 4.6 muestra los primeros diez mejores resultados según el valor de pérdida obtenido. El mínimo de estos valores es 0,5745. Tiene 5 neuronas en su primera y segunda capa intermedia. Mientras que el número de épocas es de 300, el tamaño de lote es de 32 y el optimizador usado es Adam.

**Tabla 4.6:** Simulaciones Modelos VAE: Indicador Multidimensional

Épocas	Neuronas: Capa intermedia 1	Neuronas: Capa intermedia 2	Tamaño Lote	Optimizador	Valor Pérdida
300	5	5	32	Adam	0,57463518
300	5	15	32	Adam	0,57627499
300	10	10	32	Nadam	0,57649469
300	10	5	32	RMSprop	0,57733449
300	15	15	64	RMSprop	0,57775529
300	5	15	32	Nadam	0,57805888
300	5	10	64	Nadam	0,57813044
300	10	15	32	Nadam	0,58027535
300	10	15	64	RMSprop	0,58070441
300	15	15	32	Adam	0,58173403

*Elaboración:* Los autores

## Consistencia de modelo

En la evaluación del quinto modelo se realizaron varias ejecuciones y se analizaron los valores de pérdida obtenidos en cada configuración. La tabla 4.7 muestra 15 distintas simulaciones ejecutadas para los conjuntos de entrenamiento y validación. El promedio del valor de pérdida es 0.53105 y 0.53517 para ambos conjuntos respectivamente. Así mismo, las varianzas de 0.00030 y 0.00037 indican poca variación en cuanto a los resultados del modelo. Por ende, con ello verificamos la estabilidad en la estructura de la red. En consecuencia, la arquitectura elegida del modelo final asegura que se pueda obtener la representación latente más representativa con base a los subindicadores ingresados como input.

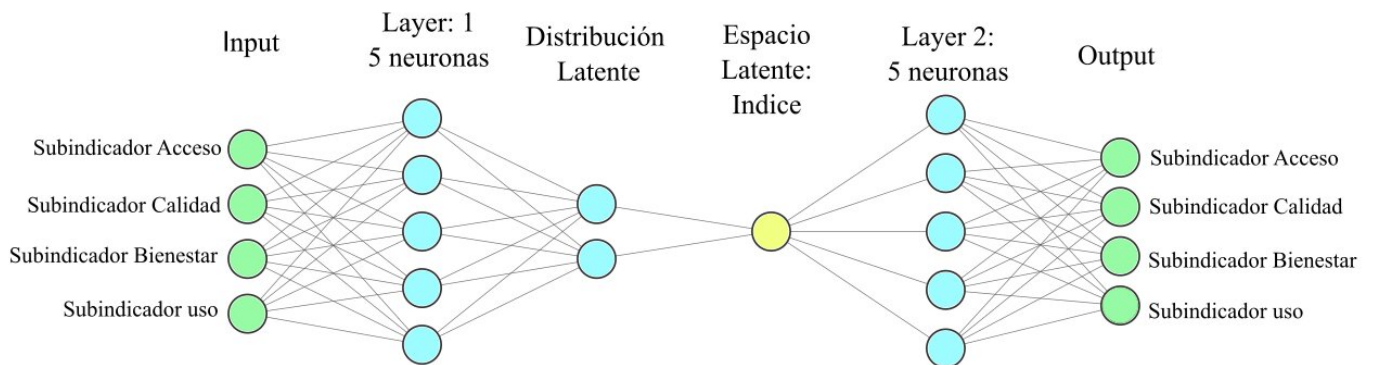
**Tabla 4.7:** Simulaciones Modelos VAE: Indicador Multidimensional

Simulación	Índice Inclusión Financiera	
	Entrenamiento	Validación
1	0,53540	0,54600
2	0,52890	0,52760
3	0,52160	0,53320
4	0,56300	0,56530
5	0,53370	0,53210
6	0,54520	0,55160
7	0,53130	0,54420
8	0,55340	0,55790
9	0,51350	0,51220
10	0,53260	0,53700
11	0,55040	0,54750
12	0,53100	0,54950
13	0,51820	0,51690
14	0,50310	0,50180
15	0,50450	0,50470
Promedio	0,53105	0,53517
Varianza	0,00030	0,00037

*Elaboración:* Los autores

Finalmente, la configuración de la red neuronal VAE final queda detallada de forma gráfica mediante la figura 4.6. Con ello, luego de haber entrenado el modelo se procede a obtener la representación latente, la cual es usada como indicador multidimensional de inclusión financiera.

**Figura 4.6:** Red Neuronal VAE: Inclusión Financiera Multidimensional



*Elaboración:* Los autores

## 4.2. Análisis Cruzado del Indicador Multidimensional de Inclusión Financiera y Variables Sociodemográficas

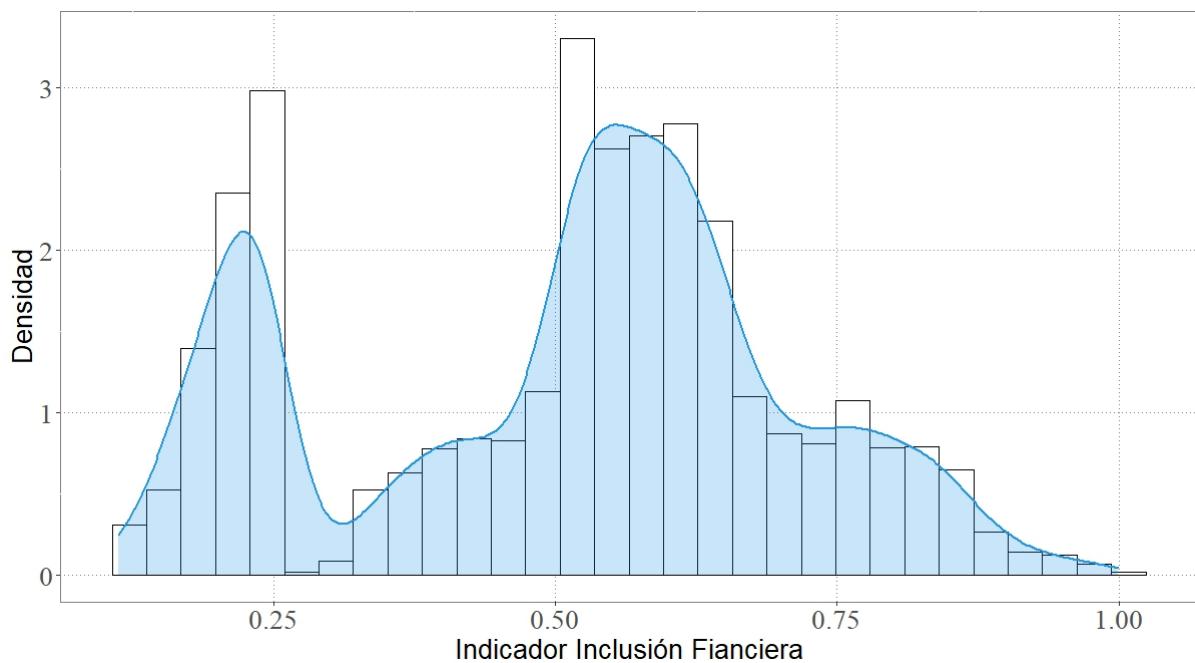
En la presente sección se realiza una descripción del indicador multidimensional de inclusión financiera. Mediante el modelamiento VAE se obtuvo la variable oculta ( $z$ ), la cual contiene para cada individuo de la muestra un valor correspondiente con respecto a la representación latente. Esta variable se escala dentro de un rango determinado con el objetivo de tener datos interpretables. Para esto se realiza una transformación lineal en donde se obtiene el valor mínimo y máximo, con base en la siguiente fórmula se obtiene el indicador de inclusión financiera multidimensional (IF):

$$IF = \frac{z - \min(z)}{\max(z) - \min(z)}$$

De esta forma, el indicador obtenido es un vector numérico cuyos valores recaen entre 0 y 1, significando que mientras más alto es el valor, mayor es el nivel de inclusión financiera.

La figura 4.7 muestra la distribución y densidad que sigue el indicador multidimensional de inclusión financiera. La forma de la distribución se asimila a una distribución multimodal y como se mencionó en el capítulo 3 la variable latente  $z$  es modelada a partir de la construcción de una distribución Gaussiana mixta, es decir, una que se forma mediante la combinación de varias distribuciones normales.

**Figura 4.7:** Distribución Indicador Multidimensional Inclusión Financiera



*Elaboración:* Los autores

Con el objetivo de determinar que el indicador es una combinación de varias distribuciones se realiza el test de profundidad de Hartigan, el cual sirve para determinar si una variable es



de tipo multimodal (Hartigan y Hartigan, 1985). El resultado del test se muestra en la tabla 4.8 e indica un *p-value* menor al nivel de confianza del 5 %, por lo tanto, existe evidencia para rechazar la hipótesis nula de unimodalidad. Esto implica que el nivel de inclusión financiera se encuentra distribuido de forma diferente según las distintas características de los individuos. Usando el método de Trang et al. (2015) se puede determinar las modas de esta distribución, las cuales son 0.2239, 0.5552 y 0.7585. La mayor frecuencia de datos se da alrededor del segundo valor, seguido del primero y por último del tercer valor.

**Tabla 4.8:** Test de Multimodalidad de Hartigan

<b>Hartigans' dip test for unimodality / multimodality</b>	
Ho: Variable es unimodal	
D	0.065452
p-value	2.2e-16
Ha: no unimodal, es decir, al menos bimodal	

*Elaboración:* Los autores

Por otro lado, para analizar la relevancia del indicador multidimensional se examinaron los niveles de inclusión financiera a nivel provincial y por área. Para ello, se calculó el promedio del indicador multidimensional y el promedio bponderado de Tukey, el cual es un estimador robusto para la media al detectar *outliers* (Mosteller y Tukey, 1977). En la tabla 4.9 se muestra los promedios para la región Costa. Se observa que para cada provincia el área rural posee los promedios más bajos, asimismo es la provincia de Esmeraldas en su área rural la que posee un menor promedio de inclusión financiera en la región costa. De igual forma, las áreas rurales de Santo Domingo, Manabí, Guayas y El Oro poseen los promedios más bajos y por debajo de 0,5. Por otro lado, el área urbana de Guayas es la que posee mayor promedio de inclusión financiera, seguido del área rural y urbana de la provincia de Santa Elena. Estas dos provincias durante 2018 se encontraron dentro de las diez provincias con menores niveles de pobreza a nivel nacional (INEC, 2018c).

**Tabla 4.9:** Indicador Promedio por Área: Región Costa

Región	Provincia	Área	Número Individuos	Promedio Aritmético	Promedio Bponderado de Tukey
Costa	El Oro	Rural	68	0,456558	0,460727
Costa	El Oro	Urbana	418	0,479918	0,486518
Costa	Esmeraldas	Rural	106	0,394638	0,311843
Costa	Esmeraldas	Urbana	111	0,457707	0,46098
Costa	Guayas	Rural	259	0,457908	0,46026
Costa	Guayas	Urbana	978	0,546969	0,555431
Costa	Los Ríos	Rural	169	0,467955	0,47189
Costa	Los Ríos	Urbana	146	0,515747	0,519632
Costa	Manabí	Rural	259	0,452273	0,459871
Costa	Manabí	Urbana	249	0,514105	0,521718
Costa	Santa Elena	Rural	36	0,531163	0,526018
Costa	Santa Elena	Urbana	80	0,563796	0,551843
Costa	Santo Domingo	Rural	52	0,391076	0,383259
Costa	Santo Domingo	Urbana	94	0,466783	0,468124

*Elaboración:* Los autores

Por otro lado, la tabla 4.10 muestra los promedios para la región Sierra, en donde son las áreas rurales de Loja, Carchi, Bolívar, Imbabura y Cotopaxi las que poseen menor promedio en inclusión financiera. Mientras que las áreas urbanas de Bolívar, Tungurahua, Cotopaxi, Azuay y Chimborazo tienen los promedios más altos de la región. De estas provincias fue Azuay quien se ubicó como la segunda provincia con menor pobreza por ingresos en 2018, mientras que fue Pichincha la primera en esta lista, la cual su promedio de inclusión financiera fue de 0.54 en el área urbana (INEC, 2018c).

**Tabla 4.10:** Indicador Promedio por Área: Región Sierra

Región	Provincia	Área	Número Individuos	Promedio Aritmético	Promedio Bponderado de Tukey
Sierra	Azuay	Rural	180	0,519357	0,52433
Sierra	Azuay	Urbana	415	0,592702	0,609377
Sierra	Bolívar	Rural	68	0,433431	0,438041
Sierra	Bolívar	Urbana	34	0,554068	0,566331
Sierra	Cañar	Rural	80	0,486229	0,486137
Sierra	Cañar	Urbana	39	0,537596	0,564444
Sierra	Carchi	Rural	51	0,422959	0,425134
Sierra	Carchi	Urbana	39	0,516819	0,520109
Sierra	Chimborazo	Rural	140	0,483389	0,486685
Sierra	Chimborazo	Urbana	76	0,573171	0,632982
Sierra	Cotopaxi	Rural	151	0,449422	0,459935
Sierra	Cotopaxi	Urbana	47	0,574124	0,587279
Sierra	Imbabura	Rural	76	0,447823	0,45415
Sierra	Imbabura	Urbana	93	0,530408	0,5647
Sierra	Loja	Rural	120	0,395381	0,386149
Sierra	Loja	Urbana	75	0,532049	0,555869
Sierra	Pichincha	Rural	190	0,478606	0,488271
Sierra	Pichincha	Urbana	790	0,531766	0,547548
Sierra	Tungurahua	Rural	176	0,514191	0,526531
Sierra	Tungurahua	Urbana	505	0,542636	0,573107

*Elaboración:* Los autores

De igual forma, la tabla 4.11 muestra los promedios de inclusión para la región Oriente. Son las áreas rurales de las provincias de Sucumbíos, Zamora Chinchipe, Morona Santiago y Orellana las que poseen los promedios más bajos de esta región. De estas provincias Morona Santiago y Sucumbíos estuvieron dentro las diez provincias con mayores niveles de pobreza por

ingresos en 2018 (INEC, 2018c). Así mismo, las áreas urbanas de Pastaza, Morona Santiago, Zamora Chinchipe y Napo son las que poseen los promedios más altos de inclusión financiera.

**Tabla 4.11:** Indicador Promedio por Área: Región Oriente

Región	Provincia	Área	Número Individuos	Promedio Aritmético	Promedio Bponderado de Tukey
Oriente	Morona Santiago	Rural	60	0,445791	0,451276
Oriente	Morona Santiago	Urbana	34	0,620157	0,655189
Oriente	Napo	Rural	38	0,519978	0,553908
Oriente	Napo	Urbana	36	0,547945	0,595632
Oriente	Orellana	Rural	47	0,4144	0,373571
Oriente	Orellana	Urbana	36	0,496296	0,49345
Oriente	Pastaza	Rural	36	0,510756	0,514714
Oriente	Pastaza	Urbana	36	0,632292	0,669637
Oriente	Sucumbíos	Rural	52	0,395703	0,392703
Oriente	Sucumbíos	Urbana	35	0,49618	0,499812
Oriente	Zamora Chinchipe	Rural	40	0,458969	0,450515
Oriente	Zamora Chinchipe	Urbana	35	0,580202	0,613525

*Elaboración:* Los autores

Por otra parte, el índice de inclusión financiera es analizado con respecto a la edad de las personas. La tabla 4.12 muestra la frecuencia del número de individuos para cinco rangos de edad de igual tamaño. El promedio de inclusión financiera muestra un nivel más bajo para las personas que se ubican en el rango de edad más alto, seguido de las personas del penúltimo rango de edad. Así mismo, los promedios más altos los tienen el tercer y segundo rango de edad respectivamente. Estos datos sugieren que los niveles de inclusión financiera tendrían un patrón de u invertida, en dónde los niveles de inclusión dejan de aumentar en cierto rango de edad. En este sentido, la gráfica 4.8 en su lado izquierdo muestra la dispersión de puntos entre edad y el

índice de inclusión financiera, así también los ejes muestran los histogramas que describen estas variables. El comportamiento de las dos variables no muestra un patrón claro con respecto su relación, sin embargo, el histograma refleja que existe un gran número de personas con un nivel de aproximadamente de 0,5 de mediana edad. También, existe un gran número de personas con niveles de aproximadamente 0.25 que se distribuyen entre todos los rangos de edad.

**Tabla 4.12:** Indicador Promedio por Rango de Edad

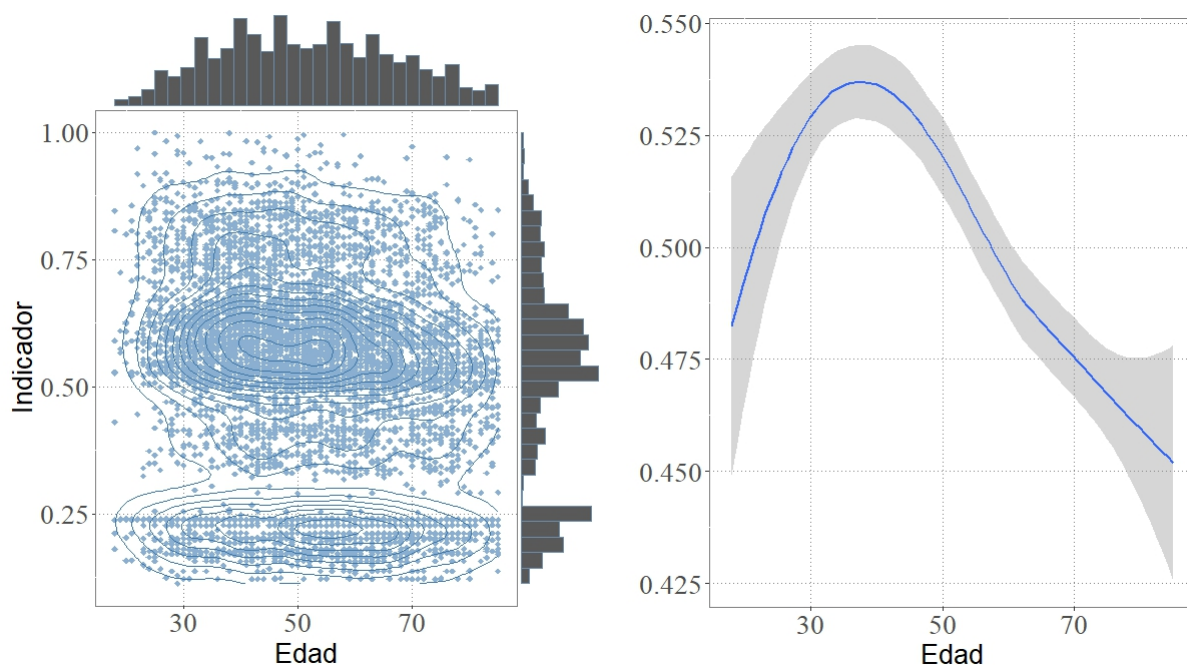
Edad	Número Individuos	Promedio Aritmético	Promedio Bponderado de Tukey
[18 - 31)	679	0,521	0,526
[31 - 44)	1801	0,532	0,544
[44 - 57)	1966	0,522	0,531
[57 - 70)	1585	0,483	0,490
[70 - 85]	895	0,470	0,476

*Elaboración:* Los autores

Con el objetivo de observar un posible patrón entre edad e inclusión financiera se usa la técnica de Loess, la cual funciona como un método no paramétrico para representar de forma gráfica una curva suavizada con respecto a dos variables (Jacoby, 2000). De este modo, la figura 4.8 muestra en el lado derecho la curva suavizada con base en su densidad de datos. Con esto se observa que existe un patrón con respecto al número de individuos. Es así que, los niveles de inclusión financiera aumentan hasta que la edad es aproximadamente 40 años, y empieza a descender desde este punto. De esta forma, se evidencia que existe una gran cantidad de individuos en el conjunto de datos que muestran este patrón.

Por otra parte, el promedio por la condición de la actividad económica del individuo se muestra en la figura 4.9 y se observa que las personas con pleno empleo son las más incluidas financieramente, seguida de las personas que se encuentran en el desempleo y la población

**Figura 4.8:** Indicador Multidimensional: Dispersión y Tendencia por Edad

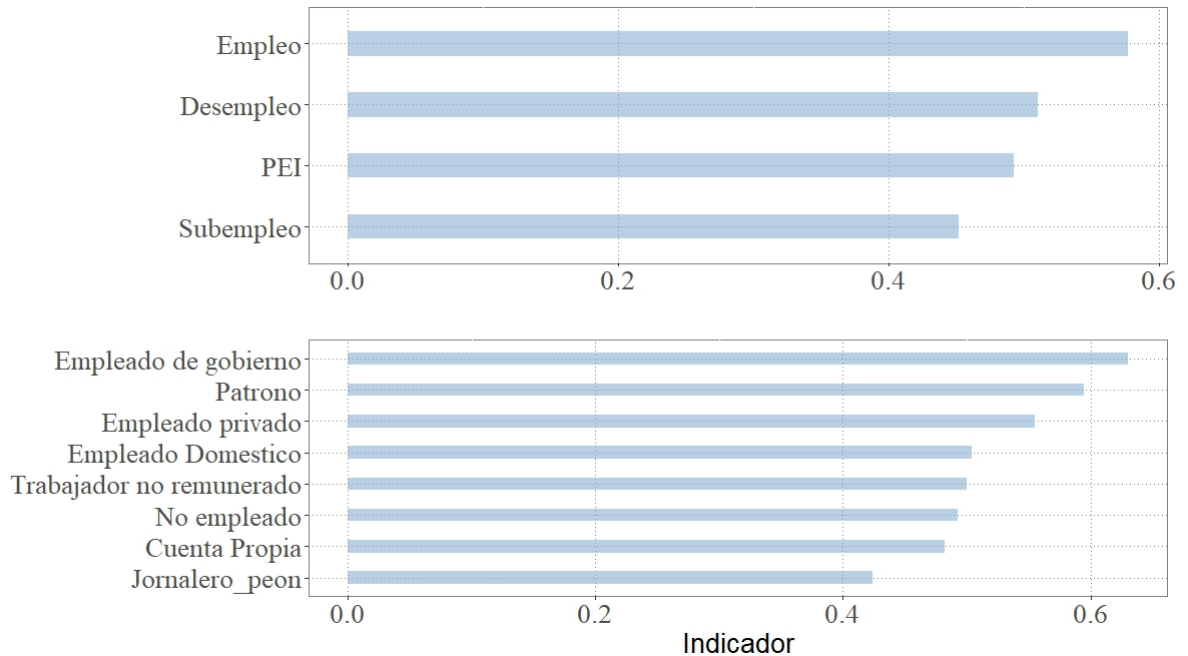


*Elaboración:* Los autores

económicamente inactiva. Mientras tanto, las personas que se encuentran en el subempleo son las menos incluidas financieramente. Este comportamiento está muy relacionado con los niveles de ingreso y por ende las personas con empleo muestran promedios más altos. Así también, los promedios en las personas que no poseen empleo pueden estar relacionados con factores no observables como el tipo de desempleo de los individuos. De igual la forma, la figura 4.9 detalla el promedio para la categoría de ocupación del individuo. Se observa que la categoría Empleado de gobierno posee el promedio más alto, seguido de las personas que son dueñas de un negocio y empleados privados. Después, se muestran promedios más bajos para las personas que no reciben un sueldo, que trabajan en la informalidad y que trabajan en el campo.

Por otro lado, la figura 4.10 muestra la distribución del indicador multidimensional con respecto al nivel de educación. El patrón muestra una media de inclusión más alta mientras más

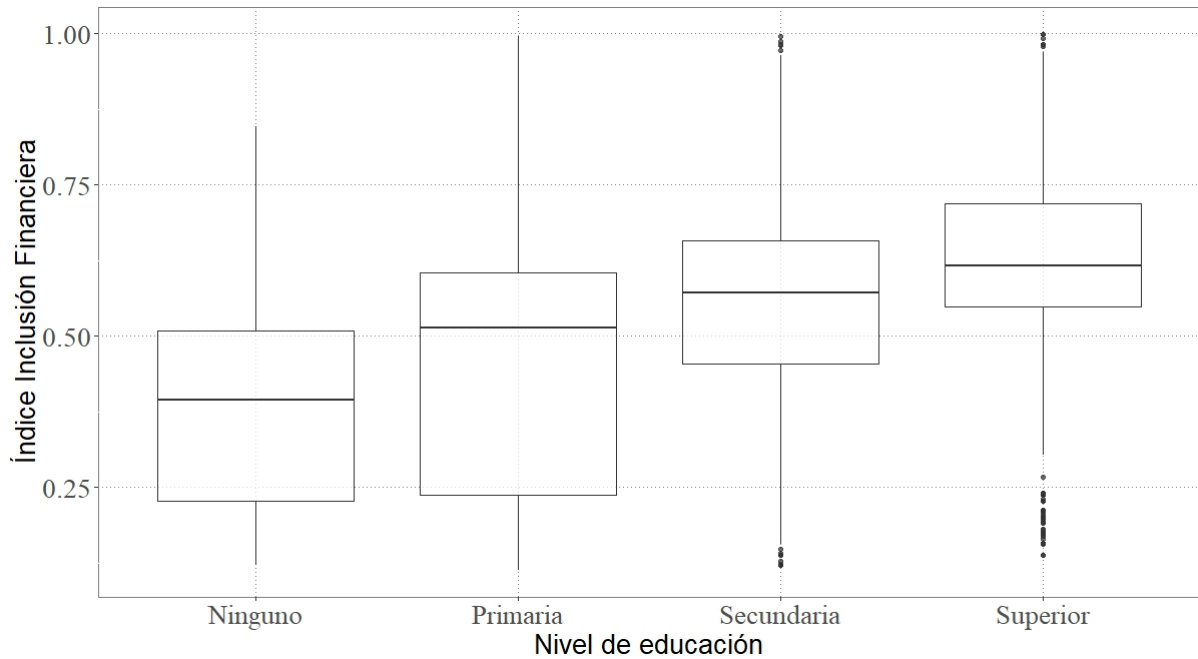
**Figura 4.9:** Indicador Multidimensional: Promedio por Categoría de Ocupación



*Elaboración:* Los autores

alto es el nivel educativo de las personas. Así mismo, se observa que existe mayor cantidad de individuos en los niveles de educación 'Ninguno' y 'Primaria'. De igual forma, los niveles educativos 'Secundaria' y 'Superior' muestran valores atípicos con respecto al nivel de inclusión financiera. Esto implica que entre los niveles de educación más altos existen individuos que tienen niveles mayores o menores con respecto a estas categorías. En todo caso, el nivel de educación presenta un patrón representativo con respecto al indicador de inclusión financiera multidimensional.

**Figura 4.10:** Indicador Multidimensional: Boxplot por Nivel de Educación

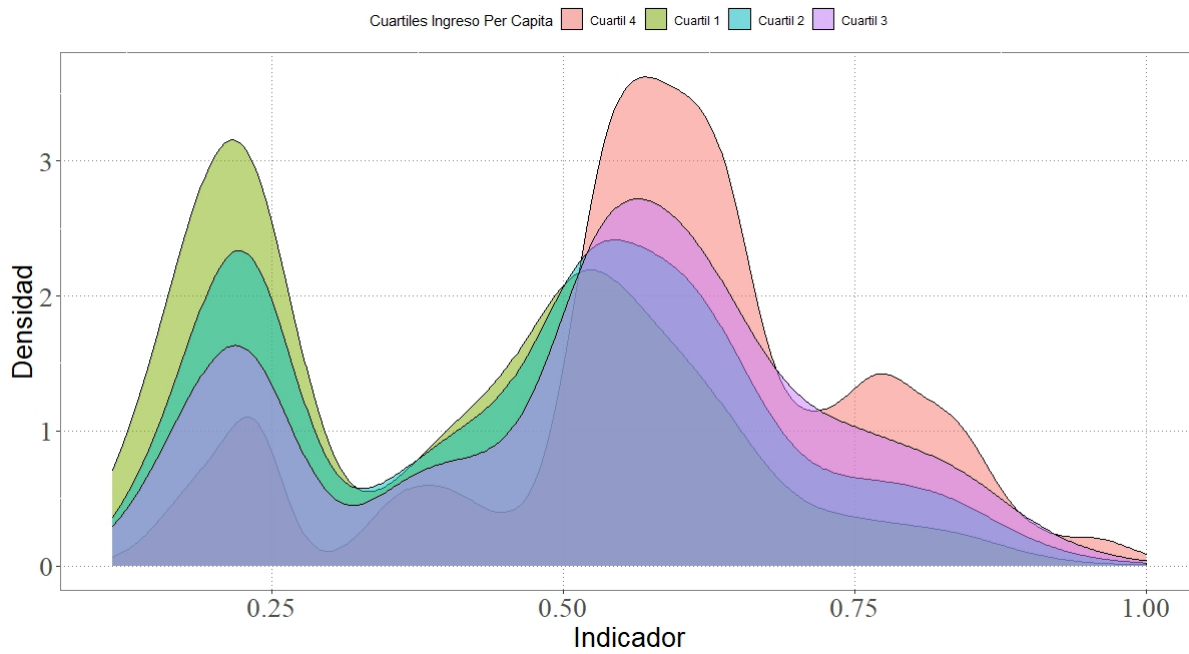


*Elaboración:* Los autores

Por último, la figura 4.11 muestra la distribución de densidad del indicador multidimensional por el cuartil de ingreso per cápita, siendo el cuartil 1 el de menores ingresos y el cuartil 4 el de mayores ingresos. La distribución refleja que los individuos del último cuartil poseen niveles de inclusión más altos con respecto a las otras categorías. El primer, segundo y tercer cuartil presentan un patrón bimodal, presentando niveles similares de inclusión financiera alrededor del valor de 0,55. Sin embargo, es en el valor menor de estas modas en que se diferencian. El primer cuartil presenta una mayor cantidad de individuos que poseen niveles de inclusión de aproximadamente 0.22, mientras que el segundo cuartil presenta un número menor de individuos sobre este valor, seguido del tercer cuartil.



**Figura 4.11:** Indicador Multidimensional: Histograma de Densidad por Ingreso Per Capita



*Elaboración:* Los autores

## 4.3. Resultados Logit

### 4.3.1. Estimación

En esta sección se presenta los resultados de la estimación de los modelos *logit*, los cuales se exponen en la tabla 4.13, además se muestra las medidas de evaluación como la sensibilidad, especificidad, precisión y la puntuación AUC (Área bajo la curva ROC que se encuentra en el Anexo 4.2.1). Las respectivas pruebas de validación para cada uno de los modelos se presentan en el Anexo D.

En el *modelo 1* se incluye únicamente las variables sociodemográficas de los individuos. En el *modelo 2* se incorpora variables de situación económica como la tenencia de vivienda propia, seguro privado y servicios básicos, sin embargo, en este modelo no se incluye la variable "área" debido a que se encuentra correlacionada con "servicios básicos". Finalmente, el modelo completo que incluye tanto variables sociodemográficas y financieras es el *modelo 3*, en el cual no se incluyó la variable "tenencia de vivienda propia" debido a que no resultó ser significativa para explicar el nivel de inclusión financiera, de igual forma se omitió la variable "seguro privado" porque al momento de incluir las variables financieras perdió su significancia. Asimismo, se omite la variable de "acceso a servicios básicos" para de esta forma poder incluir la variable "área" sin que se genere problemas de multicolinealidad. Este modelo presenta un menor valor del criterio de información Akaike (AIC) en comparación con el modelo 1 y 2.

**Tabla 4.13:** Resultado de las Estimaciones Logit de los Determinantes de la Inclusión Financiera en Ecuador

Variables	Categoría	Nivel de Inclusión Financiera		
		Modelo 1	Modelo 2	Modelo 3
Área (Ref: Urbano)	Rural	-0.212*** (0.0010)		-0.382*** (0.0000)
	Sexo (Ref: Hombre)	Mujer	0.303*** (0.0002)	0.279*** (0.0005)
Edad (Ref: Mediana edad)	Jóvenes	0.033 (0.8080)	0.048 (0.7270)	0.392** (0.0340)
	Adulto Mayor	-0.203*** (0.0050)	-0.207*** (0.0050)	-0.276*** (0.0020)
Ingrero per cápita (Ref: D Ingpc > 283)	A (Pobreza): $Ingpc < 85$	-1.254*** (0.0000)	-1.197*** (0.0000)	-0.464*** (0.0005)
	B: $85 \leq Ingpc < 163$	-0.748*** (0.0000)	-0.730*** (0.0000)	-0.225** (0.0250)
	C: $163 \leq Ingpc < 283$	-0.440*** (0.0000)	-0.440*** (0.0000)	-0.08 (0.3760)
Condición de actividad (Ref: Empleo pleno)	Subempleo	-0.455*** (0.0000)	-0.434*** (0.0000)	-0.231*** (0.0060)
	Desempleo	-0.484** (0.0310)	-0.468** (0.0360)	-0.263 (0.3480)
	PEI	-0.516*** (0.0000)	-0.478*** (0.0000)	-0.332*** (0.0040)
Estado Civil (Ref: Casado)	Separado o Divorciado	-0.530*** (0.0000)	-0.462*** (0.0000)	-0.228* (0.0550)
	Viudo	-0.405*** (0.0010)	-0.365*** (0.0030)	-0.183 (0.2040)
	Unión libre	-0.361*** (0.0000)	-0.286*** (0.0001)	-0.294*** (0.0010)
	Soltero	-0.541*** (0.0000)	-0.492*** (0.0000)	-0.066 (0.6210)
Nivel instrucción (Ref: Superior)	Ninguno	-1.705*** (0.0000)	-1.658*** (0.0000)	-1.115*** (0.0000)
	Primaria	-0.904*** (0.0000)	-0.892*** (0.0000)	-0.465*** (0.0000)
	Secundaria	-0.523*** (0.0000)	-0.510*** (0.0000)	-0.315*** (0.0030)
Miembros del hogar		0.081*** (0.0000)	0.080*** (0.0000)	0.039* (0.0530)

Etnia (Ref: Mestizo)	Indígena	-0.241** (0.0370)	-0.255** (0.0270)	-0.252* (0.0810)
	Afroecuatoriano	-0.432** (0.0260)	-0.368* (0.0600)	-0.207 (0.3960)
	Blanco	-0.173 (0.4320)	-0.172 (0.4380)	0.052 (0.8580)
Servicios básicos (Ref: No)	Si		0.485*** (0.0000)	
Vivienda propia (Ref: Si)	No		-0.002 (0.9770)	
Seguro privado (Ref: No)	Si		0.568*** (0.0002)	
Situación crediticia (Ref: Crédito aprobado)	No solicitó No puede pagar			-4.112*** (0.0000)
	Crédito rechazado			-1.920*** (0.0003)
	No solicitó No necesita			-3.733*** (0.0000)
	Crédito Informal			-1.451*** (0.0030)
Situación ahorro (Ref: Formal)	Informal			-0.879*** (0.0000)
	No Ahorra			-0.672*** (0.0000)
Califica atención (Ref: Malo)	Bueno			0.964*** (0.0000)
	No usa			-4.560*** (0.0000)
	Regular			0.394** (0.0310)
Satisfacción con los productos financieros (Ref: Insatisfecho)	No posee			-0.578** (0.0150)
	Satisfecho			0.609 (0.3150)
Constante		1.350*** (0.0000)	0.768*** (0.0000)	5.256*** (0.0000)
Sensibilidad		0.6603	0.662	0.7415
Especificidad		0.6847	0.6881	0.8229
Precisión		0.6046	0.6112	0.8089
AUC		0.7358	0.7383	0.8803
N		6,926	6,926	6,926

Log Likelihood	-4,165.48	-4,147.71	-2,836.76
Criterio de información Akaike	8,374.96	8,343.43	5,739.51
Nota:	***Significancia al nivel de 1 %.		
	** Significancia al nivel de 5 %.		
	* Significancia al nivel de 10 %.		

*Elaboración:* Los autores

Los resultados sugieren que cuando se considera únicamente las características sociodemográficas del individuo (*modelo 1*), todas las variables incluidas menos la edad categoría joven y el grupo étnico “blancos” son estadísticamente significativas y pueden explicar si el individuo se encuentra incluido financieramente o no, este modelo tiene una sensibilidad de 66 %, especificidad de 68.47 % y una precisión de 60.46 %, la puntuación AUC es de 0.7358 el cual está cercano a 1, señalando así un buen rendimiento del modelo.

En el *modelo 2*, cuando se incluye las variables referentes a la situación económica, se observa que las variables sociodemográficas mantienen su nivel de significancia para explicar la inclusión financiera, principalmente se observa que todas las categorías de ingreso son significativas con un nivel de confianza superior al 99.99 %, es decir, aun después de incluir variables que reflejan las condiciones en las que vive en individuo, como servicios básicos y tener seguro médico privado, la pobreza por ingresos tiene un alto peso para explicar la inclusión financiera. En cuanto a la información del rendimiento del modelo, se observa que los indicadores no tienen una gran variación en comparación con el *modelo 1*.

En el *modelo 3*, al incluir las variables financieras, se observa que las características sociodemográficas del individuo como el sexo, categoría de ingreso C, condición de actividad desempleo, etnia afroecuatoriana y estado civil viudo y soltero, dejan de ser estadísticamente significativos para explicar la inclusión financiera. Mientras que el resto de variables sociodemográficas siguen estando significativamente relacionadas con la inclusión financiera. El indicador de sensibilidad arroja un valor de 74 %, por tanto, este modelo tiene mejor habilidad para detectar los casos relevantes en comparación con el modelo 1 y 2. Con una especificidad de 82 %, este modelo muestra que su capacidad de discriminar los casos negativos es buena y mejor al modelo 1 y 2. Asimismo, se observa que la precisión del modelo mejora, siendo de

80.89%. Finalmente, la curva ROC muestra un valor de 88.03% señalando así que tiene una alta capacidad predictiva. Por tanto, con base en estos criterios se verifica que el *modelo 3* tiene un mejor ajuste en comparación con los *modelos 1 y 2*.

Al realizar un análisis horizontal, el resultado obtenido de la variable género resulta interesante, puesto que el signo no es el esperado en los modelos, de acuerdo con varios autores como Cámara et al. (2013), Hoyo et al. (2013) y Fungáčová y Weill (2015), el ser mujer está asociado con una menor probabilidad de que los individuos se encuentren incluidos financieramente. Además, se señala que esta variable es significativa para explicar la inclusión financiera. Sin embargo, esto no ocurre cuando se considera variables financieras. El cambio tanto de signo, como de significancia podría ser resultado del hecho de que, en los distintos estudios, se ha tomado como indicador de inclusión financiera el poseer o hacer uso de algún producto/servicio financiero, sin embargo, al no considerar simultáneamente las cuatro dimensiones de inclusión financiera el resultado de la variable mujer indicaría netamente si existe discriminación hacia un producto/servicio en específico, así como lo muestran Demirgüç-Kunt y Klapper (2012), cuyos resultados indicaron que la variable mujer solo era estadísticamente significativa y positiva en dos de sus tres modelos: tenencia de cuentas y créditos, mientras que en el modelo donde el indicador de inclusión era el ahorro formal no resultó ser significativa. Así también, Cámara et al. (2013), indicaron que la variable de género solo resultó ser estadísticamente significativa cuando analiza la inclusión financiera mediante indicadores individuales, pero en su modelo con el indicador agregado, no resultó ser significativa.

La variable de ingreso, ha mostrado que incluso después de incluir variables que se refieren específicamente a la situación financiera del individuo, encontrarse en situación de pobreza y pobreza extrema influye negativamente y es relevante para explicar la inclusión financiera considerando su significancia estadística de 99.99%. Este resultado es el esperado y se encuentra acorde a los hallazgos de autores como Anaya-Narváez y Romero-Álvarez (2018) quienes señalaron que la pobreza por ingresos disminuye significativamente la probabilidad de que los jefes de hogar se encuentren incluidos financieramente.

### 4.3.2. Efectos Marginales

Los Efectos Marginales permiten realizar la interpretación de los resultados en términos de la probabilidad de ocurrencia de que los jefes de hogar se encuentren incluidos financieramente. Por lo tanto, en la tabla 4.14 se presenta los efectos marginales del modelo obtenido, el cual considera tanto las características sociodemográficas y variables financieras del individuo.

**Tabla 4.14:** Efectos Marginales de las Estimaciones Logit de los Determinantes de la Inclusión Financiera en Ecuador

VARIABLES	Categoría	dF/dx
Área (Ref: Urbano)	Rural	-0.078*** (0.0000)
	Mujer	0.032 (0.1274)
Edad (Ref: Cat B - Mediana edad)	Jóvenes	0.088** (0.0437)
	CatC	-0.057** (0.0014)
Ingrero per cápita (Ref: D Ingpc > 283)	A (Pobreza): $Ingpc < 85$	-0.091*** (0.0001)
	B: $85 \leq Ingpc < 163$	-0.046** (0.0217)
	C: $163 \leq Ingpc < 283$	-0.017 (0.3720)
Condición de actividad (Ref: Empleo pleno)	Subempleo	-0.048*** (0.0056)
	Desempleo	-0.052 (0.3186)
Estado Civil (Ref: Casado)	PEI	-0.066*** (0.0020)
	Separado o Divorciado	-0.046** (0.0469)
	Viudo	-0.037 (0.1886)
	Unión libre	-0.06*** (0.0005)
	Soltero	-0.014

		(0.6171)
Nivel instrucción (Ref: Superior)	Ninguno	-0.183*** (0.0000)
	Primaria	-0.098*** (0.0000)
	Secundaria	-0.064*** (0.0016)
Miembros del hogar		0.008* (0.053)
Etnia (Ref: Mestizo)	Indígena	-0.05* (0.0659)
	Afroecuatoriano	-0.042 (0.3735)
	Blanco	0.011 (0.8585)
Situación crediticia (Ref: Crédito aprobado)	No solicitó No puede pagar	-0.726*** (0.0000)
	Crédito rechazado	-0.244*** (0.0000)
	No solicitó No necesita	-0.598*** (0.0000)
	Crédito Informal	-0.212*** (0.0000)
Situación ahorro (Ref: Formal)	Informal	-0.155*** (0.0000)
	No Ahorra	-0.147*** (0.0000)
Califica atención (Ref: Malo)	Bueno	0.195*** (0.0000)
	No usa	-0.567*** (0.0000)
	Regular	0.087** (0.0381)
Satisfacción con los productos financieros (Ref: Insatisfecho)	No posee	-0.132** (0.0223)
	Satisfecho	0.141 (0.346)
Nota:	***Significancia al nivel de 1 %.	
	** Significancia al nivel de 5 %.	
	* Significancia al nivel de 10 %.	

*Elaboración:* Los autores



De acuerdo con los resultados obtenidos, se puede verificar que el área de residencia donde se ubica el hogar es uno de los principales factores que determina el nivel inclusión financiera, pues los individuos que viven en el área rural tienen 7.83 % menor probabilidad de estar incluido financieramente con respecto a los que viven en el área urbana. Este resultado muestra el problema de accesibilidad física al que hace referencia Álvarez-Gamboa et al. (2021)), donde los individuos se enfrentan a barreras para acceder al sistema financiero formal por su lugar de residencia.

Al realizar un análisis por grupo de edad, se observa que si el jefe de hogar pertenece al grupo de jóvenes tiene mayor probabilidad de estar incluido financieramente en un 8.8 % con respecto a los de mediana edad, mientras que, aquellos que pertenecen al grupo de edad avanzada (superior a 59 años), disminuye la probabilidad en 5.7 %, resultado que se encuentra acorde a lo mencionado por Fungáčová y Weill (2015) referente a la existencia de un “efecto generacional”.

Con respecto al nivel de ingresos, los datos han mostrado que los individuos con menores ingresos tienen menor probabilidad de encontrarse incluido financieramente con respecto a aquellos de mayores ingresos. Específicamente, aquellos individuos cuyo ingreso se encuentra por debajo de la línea de pobreza, tienen menor probabilidad de encontrarse incluido financieramente en 9.1 % con respecto al grupo de mayores ingresos. Este resultado es el esperado, pues varios autores como Cámara y Tuesta (2014); Hoyo et al. (2013) y Tuesta et al. (2015) concluyeron que a menor nivel de ingresos existe menor probabilidad de tener acceso a productos y servicios financieros en el caso de países latinoamericanos.

Otra característica del individuo que explicaría el nivel de inclusión financiera es la condición de actividad, los individuos que se encuentran en el subempleo o pertenecen a la población económicamente inactiva tienen una menor probabilidad de estar incluido financieramente en un 4.82 % en el primer caso y en un 6.64 % en el segundo caso, con respecto a los que tienen empleo pleno, este resultado se encuentra acorde con los de Rodríguez-Raga y Rodríguez (2016)) quienes mostraron que el tener estabilidad laboral favorece al acceso a los productos y servicios financieros. Asimismo, de acuerdo con Demirgüç-Kunt et al. (2013) los individuos que se en-

cuentran fuera de la fuerza laboral tienen menor probabilidad de estar incluido financieramente.

En cuando a los resultados por estado civil, se observa que ser viudo o soltero no resulta ser significativo para explicar el nivel de inclusión financiera del individuo, sin embargo, el ser divorciado, separado o estar en unión libre, está asociado a una menor probabilidad de encontrarse incluido financieramente con respecto a los que se encuentran casados en 4.6% y 6% respectivamente, resultado que estaría acorde con autores como Al-Hussainy et al. (2008); Cano et al. (2014); Devlin (2005).

El nivel de educación del individuo resultó ser una variable relevante para explicar la inclusión financiera, la probabilidad de que los individuos que no tienen ningún nivel de estudios se encuentren incluido financieramente disminuye en 18.3% con respecto a los individuos que accedieron a la educación superior; para los que tienen estudios primarios, su probabilidad disminuye en 9.8%; mientras que los que alcanzaron el nivel secundario, la probabilidad disminuye en 6.38%. El resultado obtenido es el esperado, pues de acuerdo con varios autores como Cámara y Tuesta (2014); Tuesta et al. (2015); Rodríguez-Raga y Rodríguez (2016); Narváez et al. (2020) la educación del individuo es una de las variables determinantes más importantes a la hora de explicar su nivel de inclusión financiera, considerando que de acuerdo con Kumar (2005), esta variable podría verse relacionada con el nivel de educación financiera en el sentido de que los individuos con menor nivel de educación no tendrían los conocimientos básicos sobre finanzas, ocasionando así el uso ineficiente o autoexclusión misma del sistema financiero.

Por último, de las características sociodemográficas del individuo, la etnia indígena resultó ser significativa, indicando que la probabilidad de los individuos que pertenecen a este grupo se encuentre incluido financieramente disminuye en 5 puntos porcentuales con respecto a los que se identifican como mestizos.

Con respecto a las variables financieras, se observa que son significativas para explicar la inclusión financiera. La situación crediticia y de ahorro del individuo son estadísticamente significativas con un nivel de confianza de 99.99%, mientras que las variables referentes a calidad y bienestar tienen un nivel superior al 99%, a excepción de la categoría de bienestar donde el individuo no se encuentra satisfecho con los productos financieros la cual no tiene relevancia

estadística.

Los individuos que no solicitan crédito porque piensan que no pueden pagarlo, tienen una menor probabilidad de encontrarse incluido financieramente en un 72.58 % con respecto a aquellos que solicitaron crédito en una entidad financiera formal y fueron aprobados, este resultado confirma lo señalado por Beck et al. (2007) y Beck y Demirgüç-Kunt (2008), pues se muestra que los individuos a pesar de tener acceso a este servicio, no hacen uso del mismo por falta de ingresos, lo cual representa un problema de inclusión financiera. Así también, concuerda con lo mencionado por Demirgüç-Kunt et al. (2013) y Allen et al. (2016), quienes señalaron que la principal barrera a la inclusión financiera por el lado de la demanda es la insuficiencia de ingresos. Por otra parte, de los individuos que no solicitan crédito porque indican no necesitarlo, la probabilidad de estar incluido financieramente disminuye en 59.79 %. En cuanto a los individuos que sí solicitan crédito y no son aprobados, tienen menor probabilidad de estar incluido financieramente en un 24.39 % con respecto a aquellos que solicitan crédito en una entidad financiera formal y son aprobados. Mientras que, aquellos que solicitan crédito, pero lo hacen de manera informal (usureros, amigos, familiares), la probabilidad de estar incluido financieramente disminuye en 21.23 %, resultado que refleja precisamente las características de la inclusión financiera: tener acceso y hacer uso de los productos y servicios financieros ofrecidos por entidades financieras formales.

Por otro lado, la población que ahorra de manera informal presenta una menor probabilidad de estar incluido financieramente de 15.48 % con respecto a la población que usa estos servicios de manera formal. Además, los individuos que no ahorran tienen una menor probabilidad de encontrarse incluido financieramente en 14.71 %. Así, se observa lo mencionado por Trivelli (1997), los individuos que ahorran formalmente tienen mayor probabilidad de estar incluidos financieramente debido a que el ahorrar les permite tener mayor apertura a otros productos financieros.

La variable asociada a calidad, muestra que los individuos que califican la atención financiera como buena o regular, tienen mayor probabilidad de estar incluidos financieramente en 19.51 % y 8.69 % respectivamente con respecto a aquellos que no se encuentran satisfechos.

Mientras que, aquellos que no califican la atención debido a que no utilizan ningún punto de atención financiero (agencia bancaria, cooperativa de ahorro y crédito, cajero automático, corresponsal bancario) la probabilidad de estar incluido financieramente se ve disminuida en un 56.68 %.

Finalmente, se observa que aquellos individuos que no poseen algún producto financiero como cuentas de ahorro, corriente, depósitos a plazo o tarjeta de débito o crédito tienen menor probabilidad de encontrarse incluido financieramente en un 13.23 % con respecto a aquellos que si tienen alguno de estos productos y además se encuentran satisfechos con los servicios que le brindan estos productos.

# Capítulo V

## Conclusiones y Recomendaciones

La presente investigación ha propuesto un método que permite la medición cuantitativa del grado de inserción que un individuo posee en el sistema financiero, considerando las cuatro dimensiones teóricas de la inclusión financiera: acceso, uso, calidad y bienestar. Posteriormente, se caracterizó a los individuos de acuerdo a su nivel de inclusión financiera, considerando aspectos sociodemográficos, económicos y financieros de los individuos.

El indicador multidimensional de inclusión financiera se construyó mediante la metodología de redes neuronales *Variational Autoencoder*, la cual permitió identificar el concepto de inclusión financiera mediante la inferencia de una variable latente. En la modelización del indicador se utilizaron variables relacionadas con las dimensiones de acceso, uso, calidad y bienestar; mismas que sirvieron para construir subindicadores y que luego formaron el indicador general de inclusión financiera.

Los valores obtenidos del índice general de inclusión financiera fueron estandarizados, tomando valores entre 0 y 1. El comportamiento de este índice muestra que en el Ecuador para el año 2018 los niveles de inclusión se encuentran mayormente distribuidos en el rango de 0.55 a 0.76, seguido del rango de 0 a 0.54 y finalmente del rango de 0.77 a 1. Estos valores evidencian que un porcentaje bajo de individuos posee un mayor nivel de inclusión financiera que el resto de la población.

El índice de inclusión financiera fue categorizado de modo que cada individuo de la mues-

tra fue clasificado en dos niveles: incluidos o no incluidos financieramente. De esta forma, se determinó que en la población ecuatoriana para el año 2018, el 46.09 % de individuos jefes de hogar fueron clasificados como incluidos financieramente, mientras que el 53.91 % fueron clasificados como no incluidos financieramente.

Posteriormente, se analizó mediante modelos *logit* las características sociodemográficas, económicas y financieras de los individuos según su categorización de inclusión financiera, dando énfasis al nivel de pobreza de acuerdo a su nivel de ingresos per cápita. Los resultados señalaron que quienes se encuentran excluidos financieramente son principalmente los individuos en situación de pobreza, así como también aquellos que tienen bajos niveles de educación, lo cual a su vez estaría asociado a un bajo conocimiento sobre finanzas básicas. Se identificó que la probabilidad de que un individuo se encuentre incluido financieramente disminuye cuando posee las siguientes características: vive en el área rural, tiene bajos niveles de ingresos, está en una situación de subempleo, pertenece a la población económicamente inactiva, su estado civil es distinto a casado, tiene niveles bajos de educación y pertenece al grupo de mayor edad. Mientras que, los individuos jóvenes tienen mayor probabilidad de estar incluidos financieramente.

De manera similar, se identificó que los individuos jóvenes tienen mayor probabilidad de estar incluidos financieramente. De igual forma, se mostró que las personas que sienten mayor bienestar con los servicios financieros y los califican de mejor manera presentan una mayor probabilidad de incluirse financieramente. Por otra parte, la situación crediticia y de ahorro también resultó ser relevante para explicar la inclusión financiera. En tal sentido, cuando estas variables están asociadas al hecho de que el individuo no pueda acceder debido a un bajo nivel de ingresos, la probabilidad de estar incluido financieramente disminuye significativamente.

Los resultados observados concuerdan con los análisis empíricos de varios autores y confirman que las características de los individuos influyen en sus niveles de inclusión financiera. En este sentido, al considerar que la inclusión se relaciona con la calidad de vida y bienestar social de un individuo, es de vital importancia su análisis con el objetivo de formular políticas que faciliten el nivel de inserción al sistema financiero ecuatoriano y por ende aumentar los niveles

de inclusión financiera.

El número de personas que poseen una cuenta bancaria o ahorran son factores que dan una idea de la situación de inclusión, pero no brindan información completa y su análisis podría llevar a conclusiones equivocadas. Por consiguiente, se remarca la importancia de incorporar métodos más robustos para la medición de la inclusión financiera. De esta forma se facilita la manera de monitorear el nivel de inclusión financiera a nivel cantonal o provincial, y así poder obtener recursos informativos que permitan a los entes supervisores financieros y a las entidades financieras privadas actuar considerando las cuatro dimensiones teóricas de la inclusión financiera.

La presente investigación ha analizado los factores de la inclusión financiera y contribuye al desarrollo de la literatura existente sobre la relación de inclusión financiera y pobreza por ingresos. Se recomienda que futuras encuestas incorporen preguntas que permitan recolectar de manera objetiva información sobre las distintas dimensiones de la inclusión financiera y con ello facilitar su análisis futuro. De esta forma se superarían las limitaciones existentes con respecto a la disponibilidad de información actualizada y de calidad sobre inclusión financiera. Con ello se facilita la formulación de políticas que contribuyan a aumentar el bienestar social y, por lo tanto, la situación económica y financiera de la población ecuatoriana.

En este sentido, es imprescindible que las instituciones financieras sean capaces de brindar productos y servicios de calidad que se adapten a todos los sectores de la sociedad. Así también brindar de educación financiera, para de esta forma reducir las asimetrías de información y que la población en situación de pobreza tenga conocimiento sobre los beneficios que podrían obtener del sistema financiero formal.

# Bibliografía

- Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., ... Kiru, M. U. (2019). Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access*, 7, 158820–158846.
- Ackley, D. H., Hinton, G. E., y Sejnowski, T. J. (1985). A learning algorithm for boltzmann machines. *Cognitive science*, 9(1), 147–169.
- Agrawal, A., y Domke, J. (2021). Amortized variational inference for simple hierarchical models. *Advances in Neural Information Processing Systems*, 34, 21388–21399.
- Al-Hussainy, E., Beck, T., Demirguc-Kunt, A., y Zia, B. (2008). Household use of financial services. *World Bank Working Paper*, 502, 516.
- Alianza para la Inclusión Financiera. (2010). *La medición de la inclusión financiera para entes reguladores: Diseño e implementación de encuestas*. Descargado de [https://www.afi-global.org/sites/default/files/pdfimages/afi%20policy%20paper-SP%20\(2\).pdf](https://www.afi-global.org/sites/default/files/pdfimages/afi%20policy%20paper-SP%20(2).pdf)
- Aljalbout, E., Golkov, V., Siddiqui, Y., Strobel, M., y Cremers, D. (2018). Clustering with deep learning: Taxonomy and new methods. *arXiv preprint arXiv:1801.07648*.
- Allen, F., Demirguc-Kunt, A., Klapper, L., y Peria, M. S. M. (2016). The foundations of financial inclusion: Understanding ownership and use of formal accounts. *Journal of financial Intermediation*, 27, 1–30. doi: 10.1016/j.jfi.2015.12.003



- Álvarez-Gamboa, J., Cabrera-Barona, P., y Jácome-Estrella, H. (2021). Financial inclusion and multidimensional poverty in Ecuador: A spatial approach. *World Development Perspectives*, 22, 100311.
- Amidžic, G., Massara, M. A., y Mialou, A. (2014). *Assessing countries' financial inclusion standing-a new composite index*. International Monetary Fund.
- Anaya-Narváez, A. R., y Romero-Álvarez, Y. P. (2018). La inclusión financiera en Sincelejo (Colombia). un modelo econométrico probit. *Ecos de Economía*, 22(46), 91–110.
- Arellano, A., Cámara, N., Tuesta, D., y cols. (2014). The effect of self-confidence on financial literacy. *BBVA Research WP*, 14, 28.
- Bailey, D. L., y Thompson, D. (1990). Developing neural-network applications. *AI expert*, 5(9), 34–41.
- Banco Mundial . (2017). *Monitoring global poverty: Report of the commission on global poverty*.
- Banco Mundial . (2022). *La inclusión financiera es un elemento facilitador clave para reducir la pobreza y promover la prosperidad*. Obtenido de <https://www.bancomundial.org/es/topic/financialinclusion/overview>.
- Banco Mundial. (2018). *La inclusión financiera es un factor clave para reducir la pobreza e impulsar la prosperidad*. Descargado de <https://www.bancomundial.org/es/topic/financialinclusion/overview#1>
- Beck, T., y De la Torre, A. (2007). The basic analytics of access to financial services. *Financial markets, institutions & instruments*, 16(2), 79–117.
- Beck, T., Demirgüç-Kunt, A., y Levine, R. (2007). Finance, inequality and the poor. *Journal of economic growth*, 12(1), 27–49.

- Beck, T., y Demirgüç-Kunt, A. (2008). Access to finance: An unfinished agenda. *The world bank economic review*, 22(3), 383–396.
- Beck, T., Demirgüç-Kunt, A., y Honohan, P. (2009, 02). Access to Financial Services: Measurement, Impact, and Policies. *The World Bank Research Observer*, 24(1), 119-145. Descargado de <https://doi.org/10.1093/wbro/lkn008> doi: 10.1093/wbro/lkn008
- Berg, A. T., y Kelly, M. M. (2006). Defining intractability: comparisons among published definitions. *Epilepsia*, 47(2), 431–436.
- Blei, D. M., Kucukelbir, A., y McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518), 859–877.
- Buduma, N., Buduma, N., y Papa, J. (2022). *Fundamentals of deep learning*. ° Reilly Media, Inc.
- Cámara, N., Peña, X., Tuesta, D., y cols. (2013). Determinantes de la inclusión financiera en Perú. *Documento de trabajo*, 13, 31.
- Cámara, N., y Tuesta, D. (2014). Measuring financial inclusion: A multidimensional index. *BBVA Research Paper*(14/26).
- Cano, C. G., Esguerra, M., García, N., Rueda, L., y Velasco, A. (2014). Inclusión financiera en Colombia. *Banco de la República*, 50.
- Carrizosa, E., Restrepo, M. G., y Morales, D. R. (2021). On clustering categories of categorical predictors in generalized linear models. *Expert Systems with Applications*, 182, 115245.
- CEPAL, N. (2018). *Medición de la pobreza por ingresos: actualización metodológica y resultados*. CEPAL.
- Chakrabarty, K. (2011). Keynote address on financial inclusion. *Mumbai, September, 6*.

- Claessens, S. (2006). Access to financial services: A review of the issues and public policy objectives. *The World Bank Research Observer*, 21(2), 207–240.
- Collett, D. (2002). *Modelling binary data*. CRC press.
- Consultative Group to Assist the Poor. (2011). Global standard-setting bodies and financial inclusion for the poor: Toward proportionate standards and guidance. *White Paper prepared on behalf of the G-20's Global Partnership for Financial Inclusion*.
- Cull, R., Ehrbeck, T., y Holle, N. (2014). La inclusión financiera y el desarrollo: Pruebas recientes de su impacto. *Enfoques*, 92(1).
- Cunningham, J. P., y Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11), 1500–1509.
- Dassatti, C., y Marino, N. (2017). Construyendo un índice sintético de inclusión financiera.
- Demirgüç-Kunt, A., Klapper, L., Singer, D., y Van Oudheusden, P. (2013). The global finindex database. *Worldbank.org* <http://documents.worldbank.org/curated/en/187761468179367706/pdf/WPS7255.pdf>.
- Demirgüç-Kunt, A., y Klapper, L. F. (2012). Financial inclusion in África: an overview. *World Bank Policy Research Working Paper*(6088).
- Demirgüç-Kunt, A., Klapper, L. F., y Singer, D. (2013). Financial inclusion and legal discrimination against women: evidence from developing countries. *World Bank Policy Research Working Paper*(6416).
- Devlin, J. F. (2005). A detailed study of financial exclusion in the uk. *Journal of Consumer Policy*, 28(1), 75–108.
- Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

- Duncombe, R., y Boateng, R. (2009). Mobile phones and financial services in developing countries: a review of concepts, methods, issues, evidence and future research directions. *Third World Quarterly*, 30(7), 1237–1258.
- Ferrari, R., S.; Stengel. (2005, jan). Smooth function approximation using neural networks. *IEEE Transactions on Neural Networks* 2005-jan vol. 16 iss. 1, 16. Descargado de [libgen.li/file.php?md5=579aa1d5a565dbdd712713580e044605](http://libgen.li/file.php?md5=579aa1d5a565dbdd712713580e044605) doi: 10.1109/TNN.2004.836233
- Foster, D. (2019). *Generative deep learning: teaching machines to paint, write, compose, and play*. O'Reilly Media.
- Freund, P. G., y Rubin, M. A. (1980). Dynamics of dimensional reduction. *Physics Letters B*, 97(2), 233–235.
- Fungáčová, Z., y Weill, L. (2015). Understanding financial inclusion in china. *China Economic Review*, 34, 196–206.
- García, F. (2016). La disponibilidad y utilización de los servicios bancarios en la argentina. un estudio de sus determinantes para córdoba, entre ríos y santa fe.
- Ghorbani, A., Abid, A., y Zou, J. (2019). Interpretation of neural networks is fragile. En *Proceedings of the aai conference on artificial intelligence* (Vol. 33, pp. 3681–3688).
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning*. MIT press.
- Greenwood, J., y Jovanovic, B. (1990). Financial development, growth, and the distribution of income. *Journal of political Economy*, 98(5, Part 1), 1076–1107. Descargado de <http://www.jstor.org/stable/2937625>
- Hall, P. (1987). On kullback-leibler loss and density estimation. *The Annals of Statistics*, 1491–1519.
- Hart, S. (1989). Shapley value. En *Game theory* (pp. 210–216). Springer.

- Hartigan, J. A., y Hartigan, P. M. (1985). The dip test of unimodality. *The annals of Statistics*, 70–84.
- Hilbe, J. M. (2009). *Logistic regression models*. Chapman and hall/CRC.
- Hinkley, D. V. (1977). On power transformations to symmetry. *Biometrika*, 64(1), 121–130.
- Hosmer Jr, D. W., Lemeshow, S., y Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- Hoyo, C., Pena, X., y Tuesta, D. (2013). Factores de demanda que influyen en la inclusión financiera en México: Análisis de las barreras a partir de la enif [Working Papers]. (1336).
- Iglesias, T. (2013). Métodos de bondad de ajuste en regresión logística. *Universidad de Granada*.
- INEC. (2018a). *Encuesta nacional de empleo, desempleo y subempleo enemdu*. Quito, Ecuador.
- INEC. (2018b). *Encuesta nacional de empleo, desempleo y subempleo enemdu documento metodológico*. Quito, Ecuador.
- INEC. (2018c). *Enemdu acumulada 2018*. Quito, Ecuador.
- Jacoby, W. G. (2000). Loess:: a nonparametric, graphical tool for depicting relationships between variables. *Electoral studies*, 19(4), 577–613.
- Johnson, S., y Nino-Zarazua, M. (2011). Financial access and exclusion in kenya and uganda. *The journal of Development studies*, 47(3), 475–496.
- Kaastra, I., y Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3), 215–236.
- Katz, W., Snell, J., y Merickel, M. (1992). [29] artificial neural networks. En *Methods in enzymology* (Vol. 210, pp. 610–636). Elsevier.

- Kingma, D. P., Salimans, T., y Welling, M. (2015). Variational dropout and the local reparameterization trick. *Advances in neural information processing systems*, 28.
- Kingma, D. P., y Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- Kingma, D. P., Welling, M., y cols. (2019). An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4), 307–392.
- Klaas, J. (2019). *Machine learning for finance: principles and practice for financial insiders*. Packt Publishing Ltd.
- Kumar, A. (2005). *Access to financial services in brazil*. World Bank Publications.
- Lawrence, N. D. (2007). Learning for larger datasets with the gaussian process latent variable model. En *Artificial intelligence and statistics* (pp. 243–250).
- LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436–444.
- Levine, R. (2005). Finance and growth: theory and evidence. *Handbook of economic growth*, 1, 865–934.
- Liashchynskiy, P., y Liashchynskiy, P. (2019). Grid search, random search, genetic algorithm: a big comparison for nas. *arXiv preprint arXiv:1912.06059*.
- Lubke, G. (2018). Latent variable mixture models. En *The reviewer's guide to quantitative methods in the social sciences* (pp. 202–213). Routledge.
- Maimon, O., y Rokach, L. (2005). *Data mining and knowledge discovery handbook*.
- Maron, M. (2014). *El desarrollo del sistema financiero y su impacto en la desigualdad económica* (Tesis de Master no publicada). Universidad Iberoamericana Ciudad de México. Departamento de Economía.

- Martínez, A. (2008). El microcrédito como instrumento para el alivio de la pobreza: Ventajas y limitaciones. *Cuadernos de Desarrollo rural*, 5(61), 93–110. Descargado de <http://dx.doi.org/10.2139/ssrn.2485649>
- Masters, T. (1993). *Practical neural network recipes in c++*. Academic Press Professional, Inc.
- Meireles, M. R., Almeida, P. E., y Simões, M. G. (2003). A comprehensive review for industrial applicability of artificial neural networks. *IEEE transactions on industrial electronics*, 50(3), 585–601.
- Moreno, A., Armengol, E., Béjar Alonso, J., Belanche Muñoz, L. A., Cortés García, C. U., Gavaldà Mestre, R., ... Sánchez-Marrè, M. (1994). *Aprendizaje automático*. Edicions UPC.
- Mosteller, F., y Tukey, J. W. (1977). *Data analysis and regression. a second course in statistics. Addison-Wesley series in behavioral science: quantitative methods*.
- Narváez, A. R. A., Parra, J. B., y Alvarez, Y. P. R. (2020). Pobreza e inclusión financiera en el municipio de montería, Colombia. *Revista de ciencias sociales*, 26(1), 128–143.
- Nielsen, M. A. (2015). *Neural networks and deep learning* (Vol. 25). Determination press San Francisco, CA, USA.
- Pabón, A. M., Piñeros, H., y cols. (2006). *Situación del crédito en Colombia desde la perspectiva del sector financiero* (Inf. Téc.). Banco de la Republica de Colombia.
- Palmer, A., Montano, J. J., y Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism management*, 27(5), 781–790.
- Park, C.-Y., y Mercado, R. (2015). Financial inclusion, poverty, and income inequality in developing asia. *Asian Development Bank Economics Working Paper Series*(426).

- Patterson, J., y Gibson, A. (2017). *Deep learning: A practitioner's approach*. .<sup>o</sup>Reilly Media, Inc.
- Pérez, E., y Titelman, D. (2018). *La inclusión financiera para la inserción productiva y el papel de la banca de desarrollo*. Cepal.
- Pérez, J. M., y Martín, P. P. (2023). La curva roc. *Medicina de Familia. SEMERGEN*, 49(1), 101821.
- Raccanello, K., y Guzmán, E. H. (2014). Educación e inclusión financiera. *Revista Latinoamericana de Estudios Educativos (México)*, 44(2), 119–141.
- Rao, K. S., Baza, A. U., y cols. (2017). Barriers to access to and usage of financial services in ethiopia. *Business and Economic Research*, 7 (1), 139–149.
- Rasmussen, C. (1999). The infinite gaussian mixture model. *Advances in neural information processing systems*, 12, 554-560.
- Rhine, S. L., y Greene, W. H. (2006). The determinants of being unbanked for us immigrants. *Journal of Consumer Affairs*, 40(1), 21–40.
- Roa, M. J. (2010). *Racionalidad, uso de información y decisiones financieras* (Inf. Téc.). CIDE, División de Economía.
- Roa, M. J. (2013). Inclusión financiera en américa latina y el caribe: acceso, uso y calidad. *Boletín del CEMLA*, 59(3), 121–148.
- Roa, M. J., y Carvalho, O. A. (2018). *Inclusión financiera y el costo del uso de instrumentos financieros formales: Las experiencias de américa latina y el caribe*. Inter-American Development Bank.
- Rodríguez-Raga, S., y Rodríguez, F. F. R. (2016). Determinantes del acceso a los productos financieros en los hogares Colombianos. *Estudios Gerenciales*, 32(138), 14–24.



- Sarma, M. (2008). Index of financial inclusion. *Working Pape*(215).
- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., y Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, 20(1), 61–80.
- Sehrawat, M., y Giri, A. (2016). Financial development and poverty reduction in india: an empirical investigation. *International Journal of Social Economics*, 43(2), 106-122. Descargado de <https://doi.org/10.1108/IJSE-01-2014-0019> doi: 10.1108/IJSE-01-2014-0019
- Sharma, S., Sharma, S., y Athaiya, A. (2017). Activation functions in neural networks. *towards data science*, 6(12), 310–316.
- Shen, H., Wang, Z., Gao, C., QIN, J., YAO, F., y XU, W. (2008). Determining the number of bp neural network hidden layer units. *Journal of tianjin University of Technology*, 24(5), 13.
- Sorzano, C. O. S., Vargas, J., y Montano, A. P. (2014). A survey of dimensionality reduction techniques. *arXiv preprint arXiv:1403.2877*.
- Sotomayor, N., Talledo, J., y Wong, S. (2018). Determinantes de la inclusión financiera en el Perú: Evidencia reciente. *Superintendencia de Banca, Seguros y Administradoras Privadas de Fondos de Pensiones (SFP). Documentos de trabajo. Lima, Perú. Julio. Disponible en: [https://www.researchgate.net/publication/327288055\\_Determinantes\\_de\\_la\\_Inclusion\\_Financiera\\_en\\_el\\_Peru\\_Evidencia\\_Reciente](https://www.researchgate.net/publication/327288055_Determinantes_de_la_Inclusion_Financiera_en_el_Peru_Evidencia_Reciente)*.
- Spearman, C. (1904). Inteligencia general determinada y medida objetivamente. *Am J Psich*, 5, 201–293.
- Tran, D., Kucukelbir, A., Dieng, A. B., Rudolph, M., Liang, D., y Blei, D. M. (2016). Edward: A library for probabilistic modeling, inference, and criticism. *arXiv preprint arXiv:1610.09787*.
- Trang, N., Choisy, M., Nakagomi, T., Chinh, N., Doan, Y., Yamashiro, T., ... Anh, D. (2015). Determination of cut-off cycle threshold values in routine rt-pcr assays to assist differential

- diagnosis of norovirus in children hospitalized for acute gastroenteritis. *Epidemiology & Infection*, 143(15), 3292–3299.
- Trippi, R. R., y Turban, E. (1992). *Neural networks in finance and investing: Using artificial intelligence to improve real world performance*. McGraw-Hill, Inc.
- Trivelli, C. (1997). Crédito en los hogares rurales del Perú. *Investigaciones breves; 4*.
- Tuesta, D., Sorensen, G., Haring, A., Cámara, N., y cols. (2015). Inclusión financiera y sus determinantes: el caso argentino. *Documento de Trabajo*, 15(04).
- Van Der Maaten, L. (2009). Learning a parametric embedding by preserving local structure. En *Artificial intelligence and statistics* (pp. 384–391).
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.
- Yoon, Y., y Swales, G. (1991). Predicting stock price performance: A neural network approach. En *Proceedings of the twenty-fourth annual hawaii international conference on system sciences* (Vol. 4, pp. 156–162).
- Zaki, M. J., Meira Jr, W., y Meira, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.
- Zhang, C., Bütepage, J., Kjellström, H., y Mandt, S. (2018). Advances in variational inference. *IEEE transactions on pattern analysis and machine intelligence*, 41(8), 2008–2026.
- Zhang, X., Zhang, J., Sun, K., Yang, X., Dai, C., y Guo, Y. (2019). *Integrated multi-omics analysis using variational autoencoders: application to pan-cancer classification*. IEEE.
- Zins, A., y Weill, L. (2016). The determinants of financial inclusion in África. *Review of development finance*, 6(1), 46–57.

# Anexos

## Anexo A: Datos Perdidos

**Tabla A1:** Depuración Base de Datos

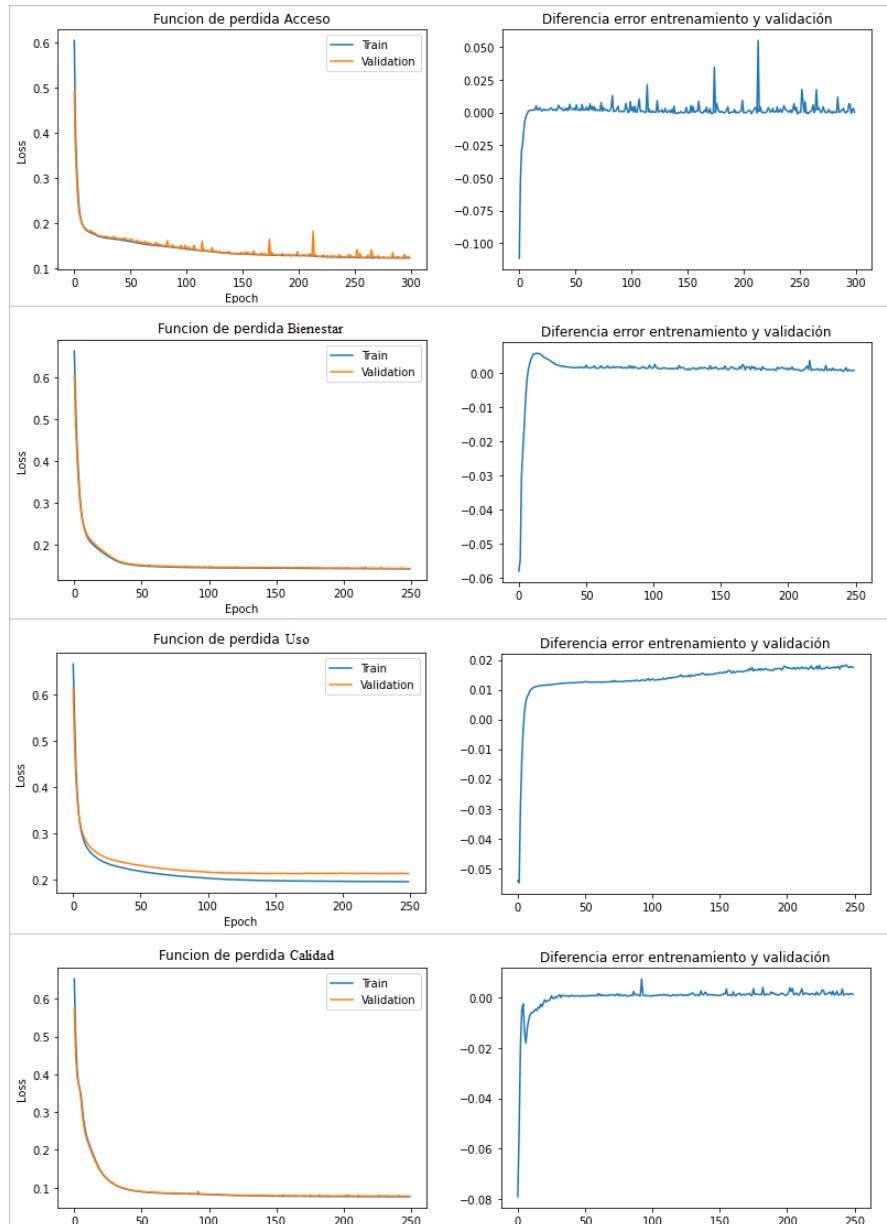
Variable / Pregunta de la encuesta	Referencia	Número Faltantes	Porcentaje
fi04	Qué transacciones realiza habitualmente - Retiro de dinero / remesas	35	0,49 %
fi14	Durante el último trimestre, Usted solicitó un crédito	2	0,03 %
pobreza	Pobreza	61	0,85 %
ingpc	Ingreso Per Cápita	3	0,04 %
p03	Edad	137	1,90 %
Provincia	Provincia	37	0,51 %
Total		275	3,82 %

*Elaboración:* Los autores



# Anexo B: Validación Modelos VAE

**Figura B1:** Funciones de Pérdida Modelos VAE



*Elaboración:* Los autores

## Anexo C: Definición Dimensiones Inclusión Financiera

Figura C1: Definiciones de Inclusión Financiera



*Fuente:* Alianza de la inclusión financiera  
*Elaboración:* Alianza de la inclusión financiera

# **Anexo D: Validación de Modelos Logit**

## **4.1. Análisis de correlación**

Notación:

**A:** Área

**B:** Califica atención

**C:** Condición de actividad

**D:** Edad

**E:** Estado Civil

**F:** Etnia

**G:** Ingreso per cápita

**H:** Nivel instrucción

**I:** Satisfacción prod. Financieros

**J:** Seguro privado

**K:** Servicios básicos

**L:** Sexo

**M:** Situación ahorro

**N:** Situación crediticia

**O:** Vivienda propia

**Tabla D1:** Matriz de coeficientes de correlación

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A	1,00	0,07	0,09	0,12	0,00	0,24	0,17	-0,54	0,25	0,05	-0,68	-0,16	-0,17	0,17	-0,35
B	0,07	1,00	0,12	0,04	0,09	0,06	0,06	-0,16	0,46	0,20	-0,14	0,01	-0,17	0,20	0,02
C	0,09	0,12	1,00	0,58	0,06	0,05	0,06	-0,32	0,26	0,32	-0,19	0,37	-0,17	0,23	-0,24
D	0,12	0,04	0,58	1,00	-0,05	0,04	-0,04	-0,34	0,14	-0,06	-0,09	0,11	-0,12	0,08	-0,35
E	0,00	0,09	0,06	-0,05	1,00	0,00	0,01	-0,07	0,16	0,18	-0,21	0,39	-0,04	0,12	0,25
F	0,24	0,06	0,05	0,04	0,00	1,00	-0,01	-0,16	0,19	0,15	-0,22	-0,03	-0,03	0,05	-0,06
G	0,17	0,06	0,06	-0,04	0,01	-0,01	1,00	-0,28	0,18	0,23	-0,10	0,00	-0,12	0,19	0,03
H	-0,54	-0,16	-0,32	-0,34	-0,07	-0,16	-0,28	1,00	-0,41	-0,38	0,49	-0,05	0,20	-0,34	0,23
I	0,25	0,46	0,26	0,14	0,16	0,19	0,18	-0,41	1,00	0,46	-0,36	0,05	-0,33	0,42	-0,05
J	0,05	0,20	0,32	-0,06	0,18	0,15	0,23	-0,38	0,46	1,00	-0,23	0,14	-0,16	0,32	0,09
K	-0,68	-0,14	-0,19	-0,09	-0,21	-0,22	-0,10	0,49	-0,36	-0,23	1,00	0,12	0,15	-0,30	0,20
L	-0,16	0,01	0,37	0,11	0,39	-0,03	0,00	-0,05	0,05	0,14	0,12	1,00	-0,05	0,10	0,03
M	-0,17	-0,17	-0,17	-0,12	-0,04	-0,03	-0,12	0,20	-0,33	-0,16	0,15	-0,05	1,00	-0,25	0,05
N	0,17	0,20	0,23	0,08	0,12	0,05	0,19	-0,34	0,42	0,32	-0,30	0,10	-0,25	1,00	-0,03
O	-0,35	0,02	-0,24	-0,35	0,25	-0,06	0,03	0,23	-0,05	0,09	0,20	0,03	0,05	-0,03	1,00

*Elaboración:* Los autores

## 4.2. Pruebas de Bondad de Ajuste

**Tabla D2:** Estadístico de la Devianza para cada modelo estimado

Devianza		
Ho: El modelo no es significativo		
Modelo	Estadístico	p-value
Modelo 1	8330.962	0.0000
Modelo 2	8295.429	0.0000
Modelo 3	5673.513	0.0000

*Elaboración:* Los autores



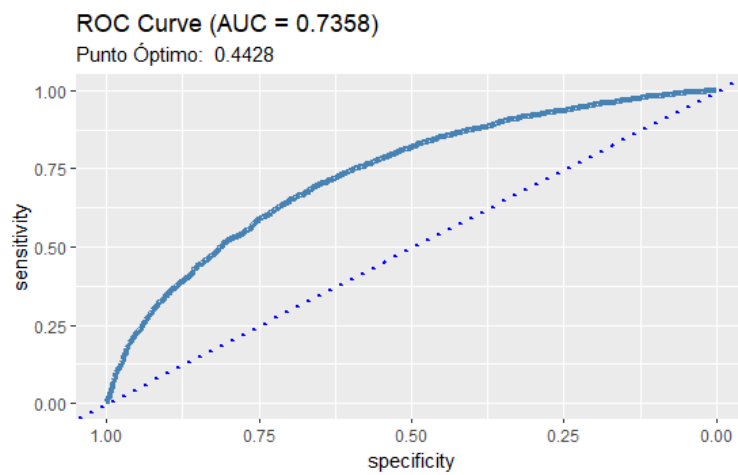
**Tabla D3:** Test Hosmer-Lemeshow Goodness of Fit "GOF" para cada modelo estimado

<b>Test Hosmer-Lemeshow Goodness of Fit</b>			
Ho: El modelo se ajusta a los datos			
<b>Modelo</b>	<b>X-squared</b>	<b>df</b>	<b>p-value</b>
Modelo 1	4.8661	8	0.7718
Modelo 2	5.0877	8	0.7482
Modelo 3	8.6576	8	0.372

*Elaboración:* Los autores

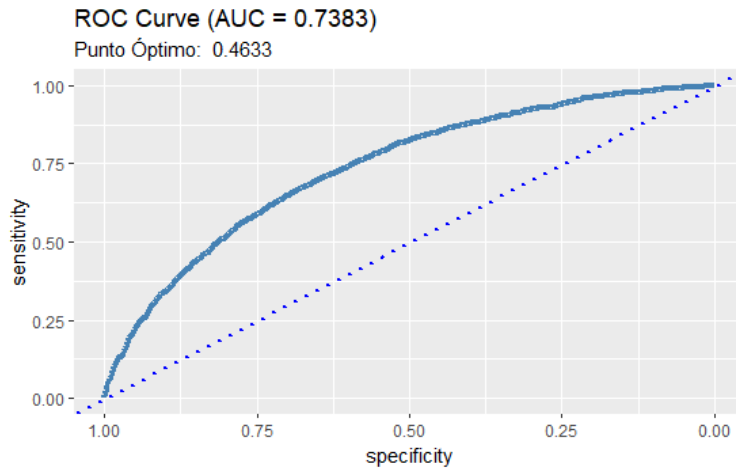
### 4.2.1. Curvas ROC

**Figura D1:** Curva ROC - Modelo 1



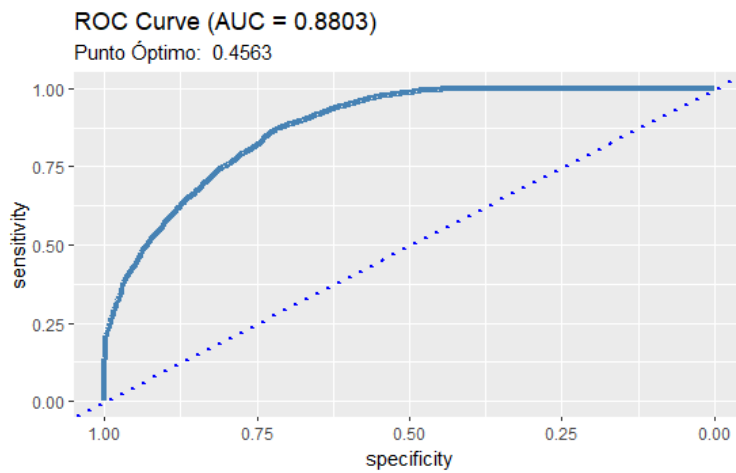
*Elaboración:* Los autores

**Figura D2:** Curva ROC - Modelo 2



Elaboración: Los autores

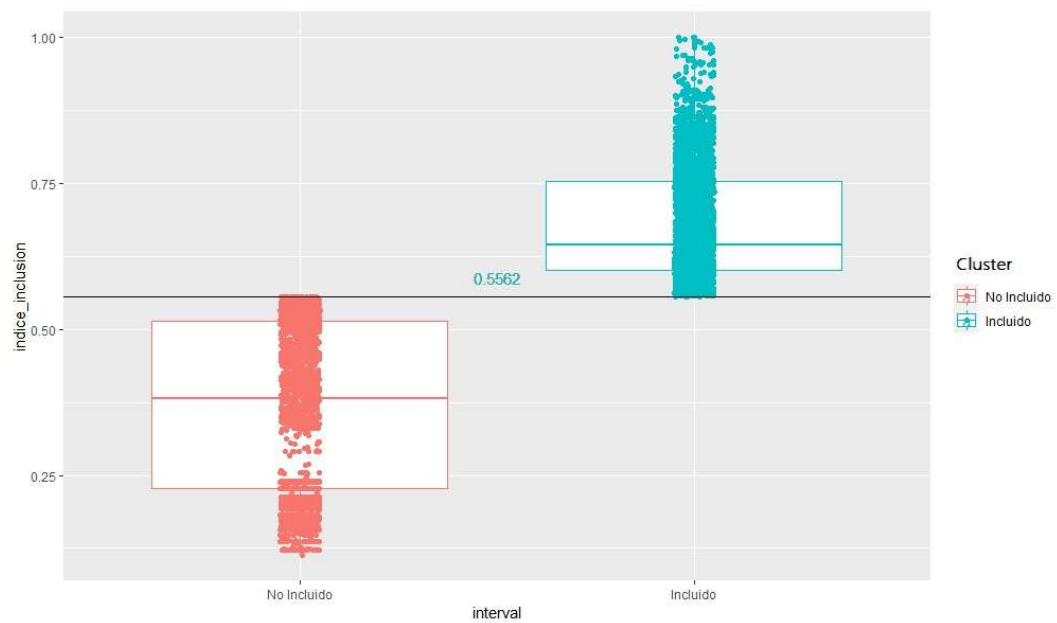
**Figura D3:** Curva ROC - Modelo 3



Elaboración: Los autores

# Anexo E: Categorización Indicador Multidimensional de Inclusión Financiera

**Figura E1:** Categorización por Clusters: Indicador Multidimensional de Inclusión Financiera



Elaboración: Los autores