



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Profesional de Estadística

Modelo de propensión a la adquisición de un producto activo o pasivo en una entidad financiera peruana

TRABAJO DE SUFICIENCIA PROFESIONAL

Para optar el Título Profesional de Licenciada en Estadística

AUTOR

Diana Milagros CÉSPEDES MALPARTIDA

ASESOR

Dr. Helfer Joel MOLINA QUIÑONES

Lima, Perú

2022



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Céspedes, D. (2022). *Modelo de propensión a la adquisición de un producto activo o pasivo en una entidad financiera peruana*. [Trabajo de suficiencia profesional de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Profesional de Estadística]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Diana Milagros Céspedes Malpartida
Tipo de documento de identidad	DNI
Número de documento de identidad	72564819
URL de ORCID	No aplica.
Datos de asesor	
Nombres y apellidos	Helfer Joel Molina Quiñones
Tipo de documento de identidad	DNI
Número de documento de identidad	40014631
URL de ORCID	https://orcid.org/0000-0003-1307-7253
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Zoraida Judith Huamán Gutiérrez
Tipo de documento	DNI
Número de documento de identidad	09890094
Miembro del jurado 1	
Nombres y apellidos	Ofelia Roque Paredes
Tipo de documento	DNI
Número de documento de identidad	06243124
Datos de investigación	
Línea de investigación	A.3.2.6. Análisis de Datos y Modelamiento de Problemas de la Sociedad (Empresas, Instituciones, Poblaciones locales, regionales y nacionales).

Grupo de investigación	No aplica.
Agencia de financiamiento	Sin financiamiento.
Ubicación geográfica de la investigación	Edificio: Universidad Nacional Mayor de San Marcos País: Perú Departamento: Lima Provincia: Lima Distrito: Cercado de Lima Calle: Av. Carlos Germán Amezaga #375 Latitud: -12.0556107 Longitud: -77.0856801
Año o rango de años en que se realizó la investigación	Abril 2022 - junio 2022
URL de disciplinas OCDE	Estadística, Probabilidades https://purl.org/pe-repo/ocde/ford#1.01.03



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

Universidad del Perú. Decana de América
FACULTAD DE CIENCIAS MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

**ACTA DE SUSTENTACIÓN DEL TRABAJO DE SUFICIENCIA PROFESIONAL EN
LA MODALIDAD VIRTUAL PARA LA OBTENCIÓN DEL TÍTULO
PROFESIONAL DE LICENCIADA EN ESTADÍSTICA
(PROGRAMA DE TITULACIÓN PROFESIONAL 2022-I)**

En la UNMSM – Ciudad Universitaria – Facultad de Ciencias Matemáticas, siendo las 09:30 horas del sábado 23 de julio del 2022, se reunieron los docentes designados como Miembros del Jurado Evaluador (PROGRAMA DE TITULACIÓN PROFESIONAL 2022-I): Dra. Zoraida Judith Huamán Gutiérrez (PRESIDENTE), Dra. Ofelia Roque Paredes (MIEMBRO) y el Dr. Helfer Joel Molina Quiñones (MIEMBRO ASESOR), para la sustentación del Trabajo de Suficiencia Profesional titulado: “**MODELO DE PROPENSIÓN A LA ADQUISICIÓN DE UN PRODUCTO ACTIVO O PASIVO EN UNA ENTIDAD FINANCIERA PERUANA**”, presentado por la señorita **Bachiller Diana Milagros Céspedes Malpartida**, para optar el Título Profesional de Licenciada en Estadística.

Luego de la exposición del Trabajo de Suficiencia Profesional, la Presidente invitó a la expositora a dar respuesta a las preguntas formuladas.

Realizada la evaluación correspondiente por los Miembros del Jurado Evaluador, la expositora mereció la aprobación de **SOBRESALIENTE**, con un calificativo promedio de **DIECIOCHO (18)**.

A continuación, los Miembros del Jurado Evaluador dan manifiesto que la participante **Bachiller Diana Milagros Céspedes Malpartida** en vista de haber aprobado la sustentación de su Trabajo de Suficiencia Profesional, será propuesta para que se le otorgue el Título Profesional de Licenciada en Estadística.

Siendo las 10:00 horas se levantó la sesión firmando para constancia la presente Acta.

Dra. Zoraida Judith Huamán Gutiérrez
PRESIDENTE

Dra. Ofelia Roque Paredes
MIEMBRO

Dr. Helfer Joel Molina Quiñones
MIEMBRO ASESOR



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú. Decana de América
FACULTAD DE CIENCIAS MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

INFORME DE EVALUACIÓN DE ORIGINALIDAD

El Director de la Escuela Profesional de Estadística, Dr. Roger Pedro Norabuena Figueroa, informa lo siguiente:

1. Operador del programa informático de similitudes: Dr. Roger Pedro Norabuena Figueroa
2. Documento evaluado: Trabajo de Suficiencia Profesional para optar el Título Profesional de Licenciada en Estadística, titulado MODELO DE PROPENSIÓN A LA ADQUISICIÓN DE UN PRODUCTO ACTIVO O PASIVO EN UNA ENTIDAD FINANCIERA PERUANA
3. Autor de la tesis: DIANA MILAGROS CÉSPEDES MALPARTIDA
4. Fecha de recepción de la tesis: 20/10/2022
5. Fecha de aplicación del programa informático de similitudes: 20/10/2022
 - Software utilizado: Turnitin
6. Configuración del programa detector de similitudes:
 - Excluye textos entrecomillados
 - Excluye bibliografía
 - Excluye cadenas menores a 40 palabras
7. Porcentaje de similitudes según programa detector de similitudes: diez por ciento (10%)
8. Fuentes originales de las similitudes encontradas:
 - Fuentes de internet: 10 %
 - Publicaciones: 0 %
9. Calificación de originalidad:
 - Documento cumple criterios de originalidad, sin observaciones

Lima, 20 de octubre del 2022



Firmado digitalmente por
NORABUENA FIGUEROA Roger
Pedro FAU 20148092282 soft
Motivo: Soy el autor del documento
Fecha: 24.10.2022 09:30:56 -05:00

Dr. Roger Pedro Norabuena Figueroa
Director

RESUMEN

La entidad financiera es una empresa peruana especializada en brindar productos financieros, sean activos o pasivos dentro del territorio nacional buscando mantener un crecimiento, y ya que existe una fuerte competencia en el mercado financiero, es que surge la necesidad de aumentar la adquisición de productos financieros en todas las regiones acorde a las características de los clientes como por ejemplo su comportamiento financiero, sus actividades económicas y su perfil demográfico, y también acorde a sus necesidades. Para resolver esta problemática se empleó información de 31,011 clientes naturales mayores de 18 años durante los meses de junio 2021 a abril 2022, todos estos clientes residentes en el Perú, los cuales fueron evaluados y admitidos en la entidad financiera, obteniendo como resultado un modelo óptimo de regresión logística, cuya buena clasificación es del 71% entre clientes que adquieren o no adquieren algún producto financiero en la entidad financiera. Además de obtener las características influyentes en el modelo como la edad, la actividad económica, el ingreso mensual, la calificación crediticia, entre otras.

Palabras clave: Análisis de Regresión Logística, Razón de probabilidades.

ABSTRACT

The financial entity is a Peruvian company specialized in providing financial products, whether assets or liabilities within the national territory seeking to maintain growth, since there is strong competition in the financial market, so the need arises to increase the acquisition of financial products in all regions according to the characteristics of the clients, such as their financial behavior, their economic activities and their demographic profile, and also according to their needs. To solve this problem, information was used from 31,011 natural clients over 18 years of age during the months of June 2021 to April 2022, all of which are clients residing in Peru, who were evaluated and admitted to the financial institution, obtaining as a result an optimal model of logistics regression, whose good classification is 71% among clients who acquire or do not acquire any financial product in the financial entity. In addition to obtaining the influential characteristics in the model such as age, economic activity, monthly income, credit rating, among others.

Keywords: Logistic Regression Analysis, Odds Ratio.

Tabla de contenido

I. INTRODUCCIÓN	8
II. DESCRIPCIÓN DE LA ACTIVIDAD	9
2.1 Reseña Breve de la Entidad Financiera	9
2.2 Organigrama Entidad Financiera	11
2.3 Problemática	12
2.4 Objetivo principal	12
2.5 Objetivos específicos	12
2.6 Breve descripción de la metodología	12
III. MARCO TEÓRICO	14
3.1 Bases Teóricas	14
Evaluación del Modelo: Estadística G	16
Razón, Odd y Odd Ratio	17
COEFICIENTE DE DETERMINACIÓN	19
DESVIANZA	19
3.2 Descripción teórica y descriptiva de la variable	20
3.3 Antecedentes	21
3.3.1 Antecedentes internacionales	21
3.3.2 Antecedentes nacionales	22
IV. METODOLOGIA	24
4.1. Métodos	24
4.2. Procedimientos para la obtención de los datos	24
4.3. Herramientas	24
4.4. Población	24
4.5. Muestra	24
V. RESULTADOS	25

5.1.1. Características demográficas	25
5.1.2. Características financieras	26
5.1.3. Porcentaje de personas que adquieren el producto.....	27
5.1.4. Adquisición del Producto Financiero según Características Demográficas	28
5.1.5. Adquisición del Producto Financiero según Características Financieras	30
5.2. Selección de variables a incluirse en los modelos	31
5.3. Ajuste de Modelo de Regresión Logística	32
5.4. Bondad de ajuste del modelo: Propensión de adquisición de un producto financiero... 	35
5.5. Matriz de Confusión.....	35
VI. CONCLUSIONES	36
VII. RECOMENDACIONES	37
VIII. BIBLIOGRAFÍA	38
IX. ANEXOS.....	39

Lista de tabla

Tabla 1 Descripción de las variables de estudio	20
Tabla 2 Tabla de frecuencia de las características demográficas de los clientes de la entidad financiera del Perú	25
Tabla 3 Tabla de frecuencia de las características financieras de los clientes de la entidad financiera del Perú	26
Tabla 4 Tabla de frecuencia de las características demográficas de los clientes de la entidad financiera del Perú según la adquisición de un producto financiero	28
Tabla 5 Tabla de frecuencia de las características financieras de los clientes de la entidad financiera del Perú según la adquisición de un producto financiero	30
Tabla 6 Análisis de variables en relación con la Adquisición de Producto Financiero	31
Tabla 7 Codificaciones de variables categóricas	32
Tabla 8 Variables en la ecuación del modelo inicial	33
Tabla 9 Prueba sobre los coeficientes del modelo	34
Tabla 10 Tabla de clasificación del modelo final de regresión logística para la propensión de adquisición de un producto financiero en la entidad financiera	35

Tabla de figuras

Figura 1 Organigrama de la Entidad Financiera Peruana	11
Figura 2 Gráfica de sector circular del porcentaje de clientes según adquisición del producto en una entidad financiera peruana entre el periodo junio 2021 y abril 2022	27

Tabla de anexos

Anexo 1 Código	39
Anexo 2 R-squared del modelo	40

I. INTRODUCCIÓN

En la situación y contexto actual de la economía en nuestro país, observamos un constante crecimiento en cuanto al comercio e industria, así como los distintos rubros y sectores de la economía.

Por ello, ante las distintas realidades y necesidades económicas de la población, ya sea para iniciar un negocio o para el consumo y obtención de recursos básicos, las instituciones financieras comenzaron a brindar diferentes opciones de productos financieros enfocándose principalmente en los empresarios rurales y urbanos.

Sin embargo, este mercado se volvió altamente competitivo, en donde los clientes podrían por diversas razones abandonar una entidad financiera.

Ante esta preocupación, el rubro de banca y finanzas se encuentran en constante seguimiento y enfoque en el cliente, en implementar modelos que permitan incrementar o mantener a los clientes y también su rentabilidad.

El presente trabajo se halla dentro de la línea de estudio de investigación del análisis de datos y modelamiento de problemas en la sociedad, ya que se utilizará información de clientes que cuentan con algún producto de crédito o préstamo en una entidad financiera del Perú.

Se planteará un modelo predictivo de regresión logística, de acuerdo a las variables o características del cliente que sean significativas en el estudio con el fin de obtener un modelo óptimo.

La primera parte del análisis consistió en el análisis descriptivo univariado y bivariado de los datos. Luego se realizó el modelamiento a los datos.

II. DESCRIPCIÓN DE LA ACTIVIDAD

2.1 Reseña Breve de la Entidad Financiera

La entidad financiera es una empresa peruana con 29 años operando a nivel nacional y especializada en brindar productos financieros, sean activos o pasivos a las micro y pequeñas empresas. Esta entidad financiera tiene como principio la inclusión financiera dentro de la realidad en la que se encuentra actualmente la sociedad peruana. El propósito es llevar acceso financiero hasta las comunidades en el interior del país.

Cuando la financiera inició, fue en el año 1994 y se estableció solo en un departamento, enfocándose principalmente en el sector rural para poder mejorar la calidad de vida de los habitantes. Posteriormente, fueron expandiéndose hacia otras regiones del Perú, estando actualmente operando en 15 regiones de la costa, sierra y selva del territorio peruano y teniendo, por ende, una evolución en el número de agencias de atención al cliente.

Los productos principales que ofrece esta financiera peruana son créditos para consumo, para medianas, pequeñas y micro empresas, agropecuario, crédito mujer (créditos especialmente otorgados a mujeres), entre otros, también en cuanto a los productos pasivos se encuentran los de ahorro corriente, ahorro CTS, ahorro inversión, para remuneraciones y de depósitos a plazo fijo.

Los objetivos de la financiera son:

Lograr y mantener un crecimiento y buen posicionamiento en el mercado financiero nacional.

Aumentar la adquisición de productos financieros en todas las regiones acorde a las características de los clientes como por ejemplo sus antecedentes, su comportamiento comercial, sus actividades económicas y su perfil financiero, y también acorde a sus necesidades.

Estas características de los clientes son importantes para la financiera porque permite tener un conocimiento acerca del cliente, y así evaluar un producto acorde a sus actividades y necesidades.

El conocimiento del cliente también permite prevenir, detectar y mitigar eventos fraudulentos que conllevarían a la entidad financiera a pérdidas rentables y además de una mala reputación dentro del mercado financiero.

2.2 Organigrama Entidad Financiera

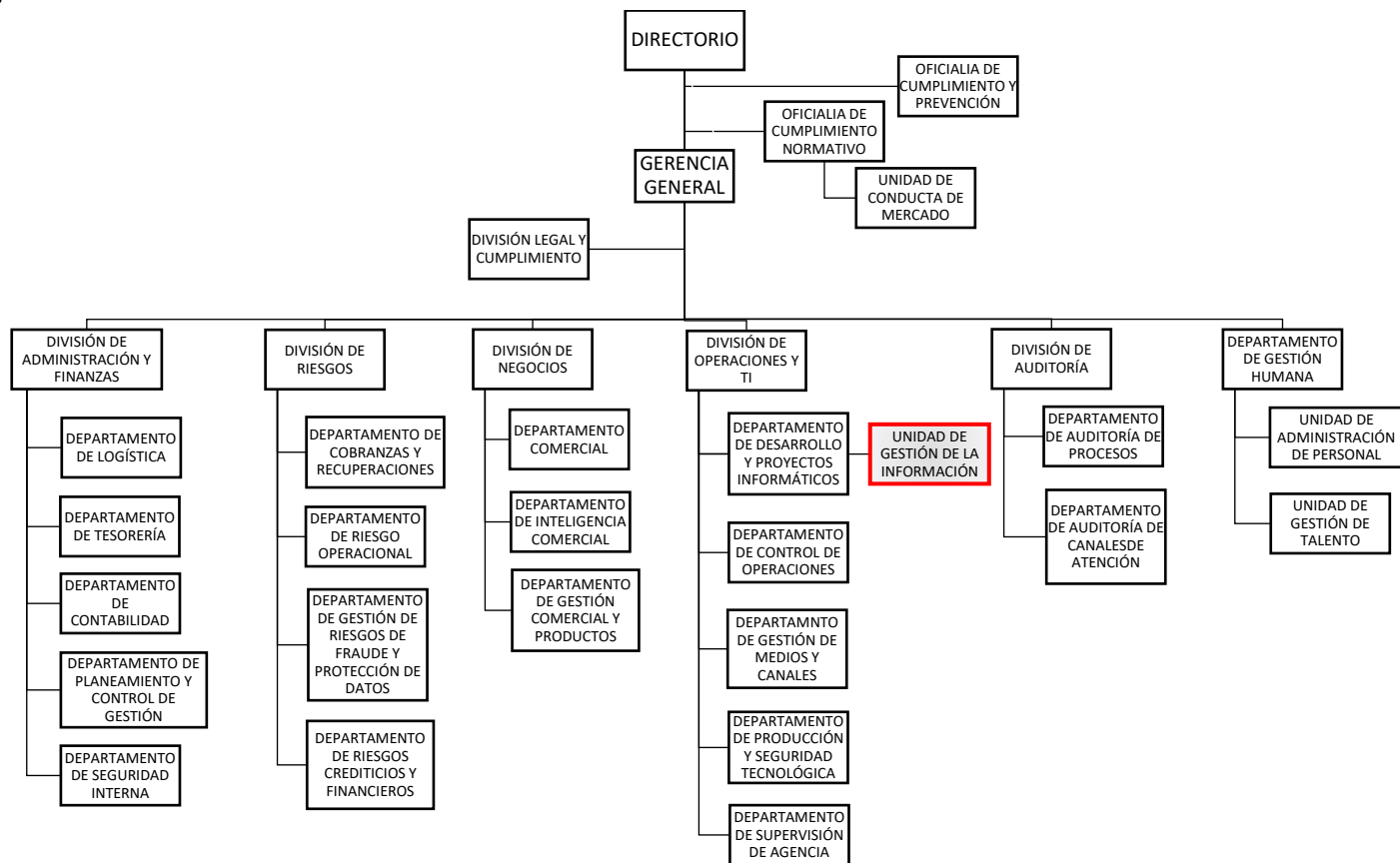


Figura 1. Organigrama de la Entidad Financiera Peruana.

Nota. Mapa conceptual de las áreas que pertenecen a la entidad financiera peruana. Extraída de la página oficial de la financiera.

2.3 Problemática

En el sector financiero muchas veces se presentan dificultades que afectan los ingresos para la empresa además de la rentabilidad financiera. Por ello es necesario proponer u ofrecer planes de acción y soluciones para las mejoras en la entidad. En el presente trabajo se aborda la problemática entorno a la adquisición de algún producto financiero. Esto dado que, en el ámbito de banca y finanzas, se está en constante monitoreo los préstamos adquiridos por los clientes con el fin de generar ingresos y mitigar el riesgo de pérdida financiera.

2.4 Objetivo principal

Obtener un modelo que pronostique la propensión de adquisición de un producto activo o pasivo en una financiera del Perú.

2.5 Objetivos específicos

Describir la relación de cada una de las variables independientes con la variable dependiente mediante la prueba de chi-cuadrado.

Medir la significancia del modelo obtenido en la adquisición de un producto financiero.

Determinar las características de los clientes más propensos a la adquisición de un producto financiero.

2.6 Breve descripción de la metodología

Para el análisis de los datos se efectuaron los siguientes pasos:

En primer lugar, la extracción de los datos se hizo mediante el sistema de gestión de base de datos SQL Server, donde se unificó y se seleccionó las variables a utilizar en el estudio.

Para el tratamiento de los datos se utilizó el software libre Rstudio y el programa estadístico SPSS.

Se definió la población de estudio como todos los clientes naturales de la entidad financiera durante los meses de junio 2021 a abril 2022.

Descripción del tipo y diseño de investigación y la definición de las variables utilizadas en el estudio.

Un análisis descriptivo univariado y bivariado de la data.

Generación del modelo logístico y prueba.

Evaluación de resultados.

III. MARCO TEÓRICO

3.1 Bases Teóricas

Camarero, A., Almazán, A. y Mañas, B. (2013), definen la regresión logística como un método de análisis, el cual permite relacionar una variable de respuesta dicotómica con un conjunto de variables predictoras. La regresión logística puede considerarse una extensión de los modelos de regresión lineal, con la particularidad de que el dominio de salida de la función está acotado al intervalo $[0,1]$ y que el método de estimación aplicado es mediante la función de máximo-verosimilitud.

Hosmer, D. & Lemeshow, S. (2013), en el modelo de regresión logística consideran lo siguiente:

Se tiene un conjunto de k variables independientes denotadas por el vector $X = (x_1, x_2, \dots, x_p)$ donde la probabilidad condicional de que el resultado esté presente se denota por $\Pr(Y = 1 | X) = \pi(X)$.

Entonces la forma específica del modelo de regresión logística es:

$$P(Y=1|X_1=x_1, X_2=x_2, \dots, X_p=x_p) = \pi(X) = \frac{e^{g(X)}}{1+e^{g(X)}}$$

donde

$$g(x) = \ln \left(\frac{\pi(X)}{1-\pi(X)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

La importancia de esta transformación es que $g(x)$ tiene muchas de las propiedades deseables de un modelo de regresión lineal. El logit, $g(x)$, es lineal en sus parámetros, puede ser continuo y puede variar de $-\infty$ a $+\infty$, dependiendo del rango de x .

Ajustar el modelo de regresión logística a un grupo de datos requiere que estimemos los valores de $\beta_0, \beta_1, \beta_2, \dots, \beta_p$.

Como se asume que las observaciones son independientes, la función de probabilidad se obtiene como el producto de los términos de la siguiente manera:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

De esta manera, la función de máxima verosimilitud para estimar los parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ que maximiza $l(\beta)$, se define como:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

En la cual, calculando el logaritmo de la función de máxima verosimilitud, se tiene:

$$L(\beta) = \ln [l(\beta)] = \sum_{i=1}^n \{y_i \ln [\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)]\}$$

Para encontrar el valor de β que maximiza $L(\beta)$, diferenciamos $L(\beta)$ con respecto a β_0 y β_1 y establecemos las expresiones resultantes en cero.

Estas ecuaciones, conocidas como las ecuaciones de probabilidad, son:

$$\sum [y_i - \pi(x_i)] = 0 \quad \text{y} \quad \sum x_i [y_i - \pi(x_i)] = 0$$

Donde i varía de 1 a n .

Estas ecuaciones no son lineales en β_0 y β_1 , y para su solución requieren de métodos iterativos computacionales que pueden ser encontrados en softwares estadísticos, por ejemplo, estas ecuaciones se pueden obtener usando un procedimiento iterativo de mínimos cuadrados ponderados.

Evaluación del Modelo: Estadística G

Para determinar la importancia y significancia del modelo, es decir evaluar si las variables independientes son significativas o no, se realizan las siguientes hipótesis:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

$$H_1: \text{Por lo menos un } \beta_i \neq 0, \forall i = 1, 2, 3, \dots, k$$

con un nivel de significación.

El estadístico G, el cual se plantea para la evaluación de la significancia del modelo es la diferencia del valor de la desviación del modelo sólo con la constante y del valor de la desviación del modelo donde se incluye a las variables independientes, este estadístico sigue una distribución Chi – Cuadrada con k grados de libertad, teniéndose:

$$G = D(\text{modelo sin variables}) - D(\text{modelo con variables})$$

$$G = -2 \ln \left(\frac{\text{verosimilitud del modelo sin variables}}{\text{verosimilitud del modelo con variables}} \right)$$

Donde:

$$-2 \ln(\text{Verosimilitud del modelo sin variables}) = -2n_1 \ln(n_1) + n_0 \ln(n_0)$$

$$- n \ln(n)$$

$$-2 \ln(\text{Verosimilitud del modelo con variables})$$

$$= -2 \sum_{i=1}^n y_i \ln(\hat{p}_i) + (1 - y_i) \ln(1 - \hat{p}_i)$$

Luego la estadística G tiene una distribución Chi-cuadrado con k grados de libertad (g.l.), bajo la hipótesis nula. Se rechaza H_0 a un nivel de significancia α , si:

$$G > \chi^2_{k,\alpha}$$

Se concluye que por lo menos uno de los parámetros es diferente de cero.

Razón, Odd y Odd Ratio

Una razón o ratio es el cociente entre dos cantidades, la cual refleja cuantas veces una cantidad es mayor o menor respecto a la otra.

El término Odd en inglés se refiere a la razón que se establece entre la ocurrencia -o su probabilidad- de un suceso respecto a su no ocurrencia.

$$Odd = \frac{p}{q} = \frac{p}{1-p}$$

Y mediante dicha definición la relación entre Odd y p podemos expresarla de la siguiente forma:

$$p = \frac{Odd}{1 + Odd}$$

A partir del Odd se puede definir el logit como el logaritmo del Odd.

Así:

$$Logit = Ln(Odd) = Ln\left(\frac{p}{1-p}\right)$$

Los Odd Ratio, también conocido como OR, puede interpretarse como una razón de probabilidades.

$$Odd Ratio = OR = \frac{Odd_A}{Odd_B} = \frac{\frac{p_A}{(1-p_A)}}{\frac{p_B}{(1-p_B)}}$$

Según Garson, D. (2014), la regresión logística se centra en los siguientes términos:

- Odds: En la regresión logística binaria, las Odds son la probabilidad de obtener un "1" dividido por la probabilidad de obtener un "0". Es decir, se predice "1" y generalmente "0" es la categoría de referencia.
- Odd Ratio: Son la principal medida del tamaño del efecto para la regresión logística, cuanto más lejos se encuentre de 1 en cualquier dirección, mayor será el efecto.

Interpretación de los parámetros:

La interpretación se realizará a partir de los Odds Ratio (OR), el cual es el cociente de probabilidades entre π si ocurre el suceso y $(1 - \pi)$ si no ocurre el suceso.

$$OR = \frac{\pi}{1 - \pi} = e^{X'\beta}$$

donde:

OR=1, indica que la variable explicativa es independiente con el evento.

OR<1, indica que la variable explicativa es un factor protector.

OR>1, indica que la variable explicativa es un factor que influye.

- Logit: Es la función utilizada en la regresión logística para transformar la variable dependiente antes de intentar predecirla.
- Log odds: Las probabilidades de registro son el coeficiente predicho por la regresión logística y se llama "logit". Es el registro natural de las probabilidades de la variable dependiente que iguala algún valor (por ejemplo, 1 en lugar de 0 en la regresión logística binaria). Por lo tanto, las probabilidades del registro son iguales al registro natural de la probabilidad de que el evento ocurra dividido por la probabilidad de que el evento no ocurra:

$$\ln(\text{prob}(\text{evento})) = \ln\left(\frac{\text{prob}(\text{evento})}{\text{prob}(\text{noevento})}\right)$$

COEFICIENTE DE DETERMINACIÓN

- **R² de Cox y Snell:** Este coeficiente se utiliza para estimar la proporción de varianza de la variable dependiente explicada por las variables predictoras (independientes). Sus valores oscilan entre 0 y 1, de tal forma que 0 indicará un efecto muy bajo de las variables independientes, mientras que en la proximidad de 1 mostrará un efecto considerable.
- **R² de Nagelkerke:** Es una versión corregida de la R² de Cox y Snell, dado que este último tiene un valor máximo inferior a 1, incluso para un modelo “perfecto”. El R² de Nagelkerke corrige la escala del estadístico para cubrir el rango completo de 0 a 1.

DESVIANZA

O también $2\log$ de la verosimilitud, mide el correcto ajuste del modelo a los datos, y en tanto más pequeño sea dicho valor, mejor será el ajuste del modelo.

3.2 Descripción teórica y descriptiva de la variable

Tabla 1.

Descripción de las variables de estudio.

Variable	Descripción	Tipo	Escala
Sexo	Género del cliente	Cualitativa	Nominal
Estado Civil	Estado de relación de familia en el que se encuentra el cliente.	Cualitativa	Nominal
Edad	Edad del cliente	Cuantitativa	Numérica
Rango Monto Ingreso	Rango del monto de ingreso mensual del cliente	Cualitativa	Ordinal
Tipo Ocupación	Tipo de ocupación del cliente: Dependiente, Independiente u Otro.	Cualitativa	Nominal
Condición de Vivienda	Condición de la vivienda del cliente: Propia, de familiares o alquilada.	Cualitativa	Nominal
Indicador Cliente Deudor	Actividad económica a la que se dedica el cliente.	Cualitativa	Nominal
Antigüedad (meses)	Número de meses desde que el cliente ingresó a la entidad financiera.	Cuantitativa	Numérica
Calificación Crediticia	Calificación Crediticia del cliente de la Entidad Reguladora (SBS)	Cualitativa	Ordinal
FlagAdquisicionProducto	Cliente mayor de 18 años que haya adquirido algún producto financiero en la entidad durante el periodo de referencia junio 2021 a abril 2022 (últimos 11 meses).	Cualitativa	Nominal

3.3 Antecedentes

3.3.1 Antecedentes internacionales

Ayala (2020) en su tesis titulada, “Modelo de Propensión a la toma de Créditos de Consumo en una Empresa del Sector Financiero”, para optar el grado de magíster en Data Science tiene como objetivo obtener un modelo predictivo que permita conocer si un cliente adquirirá o no un producto crediticio de consumo en una cooperativa de ahorros y créditos del país. Se emplearon datos de clientes nacionales en un periodo de 20 meses. El interés de este estudio es poder identificar las variables que inciden en el comportamiento crediticio de cada persona al momento de tomar un crédito. En esta tesis se emplearon técnicas de machine learning para predecir la propensión a la compra del crédito en los siguientes meses en la empresa, donde se utilizó el algoritmo XGBoost el cual permitió aumentar la precisión del modelo obtenido conforme a cada iteración sin caer en el overfitting o sobreajuste. Se concluye que las variables más influyentes y con mayor importancia para el modelo de toma de créditos corresponde al nivel de ingreso, la edad, al nivel de endeudamiento, entre otras.

Esguerra, y Niño (2017) en su proyecto de investigación titulado, “Propensión de los clientes de tarjeta de crédito del banco de Bogotá hacia la educación financiera”, para optar el grado académico de Especialización en Finanzas tiene como finalidad de determinar la propensión de compra de tarjetas de crédito en el Banco de Bogotá. En esta investigación se empleó una encuesta para la recolección de información a 50 personas, además se empleó una metodología de investigación analítica descriptiva, tanto cualitativo como cuantitativo. El interés de este estudio es generar una propuesta para que el banco centre en brindar una oportuna información financiera al cliente antes de realizar la colocación del crédito para así evitar futuros impagos o cancelaciones. En este estudio se concluye que a medida que el cliente tenga más antigüedad en el uso de tarjetas de crédito tiene más conocimiento de como emplear su crédito y por ende mayor propensión a la adquisición de nuevos créditos a futuro.

Gross (2018) en su tesis titulada “Elaboración de un modelo de propensión al crédito de consumo capaz de discriminar a nivel de individuo utilizando solamente información financiera de carácter pública”, para optar el grafo de magíster en Ciencias de la Ingeniería tiene como objetivo generar un modelo valido que permita detectar las necesidades crediticias de los clientes del sistema financiero, además de medir la propensión de compra de créditos de consumo. Para este fin se empleó información pública financiera de 85 millones de registros pertenecientes a 1.6 millones de personas. Se realizó la selección de variables relevantes para el modelo final evaluándolo en el dataset de test, se comparó los modelos de regresión logística, random forest, máquina de soporte vectorial, redes neuronales y boosting, donde se obtuvo que el mejor modelo se generó mediante el algoritmo XGBoost, y donde las variables relevantes en el modelo fueron la edad, si el cliente es deudor, así como su estrato socioeconómico, entre otras.

3.3.2 Antecedentes nacionales

Chipana (2018) en su tesis titulada “Evaluación de la gestión crediticia en el otorgamiento de créditos en la Financiera Credinka Agencia Yunguyo. Periodo 2015 – 2016”, para optar el título profesional de Contador Público tiene como finalidad analizar la gestión crediticia respecto a la recuperación de créditos en la entidad financiera Credinka. Se empleó un tipo de investigación descriptivo y analítico, además se emplearon dos cuestionarios para la obtención de información. Se obtuvo que, las variables influyentes fueron la antigüedad, la calificación crediticia, su nivel de endeudamiento, la actividad económica del cliente, entre otras características. También se obtuvo que conforme mayor es el monto del crédito desembolsado entonces existe mayor riesgo de que el cliente incumpla con el pago aumentando así el índice de morosidad. Teniendo así que los créditos financieros con montos altos son los créditos menos recuperados.

Barrueta, y Castillo (2018) en su tesis titulada “Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos” tiene como objetivo proponer un modelo predictivo para generar una mejora en la toma de decisiones respecto a la retención y fidelización de los clientes con mayor tendencia al abandono o deserción crediticia bancaria. Se empleó el análisis utilizando la herramienta SAP Predictive Analytics el cual permitió identificar las relaciones entre las variables de estudio. Se trabajó con información de 20mil clientes clasificados entre Desertores y No Desertores. El interés de este estudio fue cuantificar la deserción de clientes en los bancos peruanos, obteniendo así un modelo cuya buena clasificación fue de 93.2%, con lo que contribuiría a poder aplicar el modelo dentro del entorno organizacional y así predecir el comportamiento de clientes respecto a su actividad financiera. Se concluyó que las variables influyentes en el modelo propuesto fueron antigüedad, edad del cliente, nivel de calificación de riesgo, el ingreso mensual, el número de transacciones, entre otras en un modelo con capacidad de alcance de una buena precisión en la medición y predicción en la deserción de clientes.

Zarabia (2020) en su tesis titulada “Identificación de la propensión a la adquisición de un subproducto de una tarjeta de crédito en una entidad bancaria” tiene como objetivo identificar a los clientes con tarjeta de crédito de una entidad financiera más propensos a la adquisición del subproducto llamado Extra Línea. Se empleó información histórica de clientes durante los 11 últimos meses, de febrero a diciembre del 2016. Para el desarrollo de este análisis se utilizó la metodología CRISP, usando para el desarrollo los algoritmos de Regresión Logística, Árboles de Decisión y Random Forest. El interés de este estudio es identificar factores que influyen en la propensión a responder positivamente a la adquisición del subproducto de interés durante la campaña. Se determinó que las variables principales influyentes en la adquisición de una extra línea, fueron principalmente la antigüedad del cliente en la financiera, el número de productos en el sistema financiero, monto de su línea de crédito del cliente, deuda en su tarjeta de crédito del cliente, entre otras. Además de obtener una tasa de buena clasificación de 67.4% en la adquisición del subproducto.

IV. METODOLOGIA

Para el presente trabajo se efectuó una serie de pasos que fueron descritos en el capítulo III.

4.1. Métodos

El trabajo es de tipo correlacional puesto que se desea ver qué relación existe en la adquisición de un producto financiero ya sea activo o pasivo en la financiera y las características del cliente que lo adquiere. El diseño de la investigación es no experimental, puesto que se realizó sin manipular premeditadamente las variables y también es transversal puesto que este diseño es una investigación observacional, que mide una o más características (variables), en un momento dado.

4.2. Procedimientos para la obtención de los datos

Para realizar este trabajo se tuvo que consolidar los datos de clientes registrados por la financiera con los datos de productos activos desembolsados y productos pasivos aperturados. Estos datos se encuentran en el sistema de gestión de base de datos SQL Server, donde se unificó y se seleccionó las variables a utilizar en el estudio. Esta tabla final fue utilizada para hacer el análisis de modelamiento.

4.3. Herramientas

Para este trabajo se empleó el programa estadístico SPSS y el software libre R.

4.4. Población

31,011 clientes naturales mayores de 18 años de una entidad financiera en el Perú durante los meses de junio 2021 a abril 2022.

4.5. Muestra

No se seleccionó una muestra.

V. RESULTADOS

5.1. Análisis descriptivo de las características demográficas y financieras de los clientes

5.1.1. Características demográficas

Tabla 2.

Tabla de frecuencia de las características demográficas de los clientes de la entidad financiera del Perú.

Características	Frecuencia	Porcentaje	Media ± DE
Sexo			
Masculino	15406	49.7%	
Femenino	15605	50.3%	
Estado Civil			
Soltero o Sin Pareja	27697	89.3%	
Casado	3314	10.7%	
Edad			35.75 ± 12.817
Rango Monto Ingreso			
Hasta 1025	6570	21.2%	
1025 a 2500	14610	47.1%	
2500 a 4000	5161	16.6%	
4000 a 7000	3197	10.3%	
7000 a más	1473	4.7%	
Tipo Ocupación			
Ocupación Dependiente	4270	13.8%	
Ocupación Independiente	21980	70.9%	
Otros	4761	15.4%	
Condición de Vivienda			
Alquilada	3598	11.6%	
De Familiares	16357	52.7%	
Propia	11056	35.7%	

En la Tabla 2, en cuanto a las características demográficas de los clientes evaluados y admitidos en la entidad financiera del Perú, el género predominante fue el sexo femenino siendo el 50.3% del total de clientes, durante el periodo de junio 2021 a abril 2022.

También, se obtuvo que el 89.3% de los clientes era soltera o no se encontraban con pareja. Donde, además, la edad promedio de clientes es de 36 años.

Se observó también que el mayor volumen de clientes evaluados y admitidos en la entidad financiera peruana tienen su monto de ingreso mensual en el rango de 1025 a 2500 soles, siendo este el 47.1%.

Se obtuvo también que el 70.9% de los clientes cuentan con un tipo de ocupación independiente, por otro lado, el 52.7% de los clientes evaluados y admitidos en la entidad financiera viven en una vivienda de familiares.

5.1.2. Características financieras

Tabla 3.

Tabla de frecuencia de las características financieras de los clientes de la entidad financiera del Perú.

Características	Frecuencia	Porcentaje	Media ± DE
Indicador Cliente Deudor			
Cliente No Deudor	29028	93.6%	
Cliente Deudor	1983	6.4%	
Antigüedad (meses)			5.10 ± 3.082
Calificación Crediticia			
Normal	29982	96.7%	
Con problema potencial	443	1.4%	
Deficiente	242	0.8%	
Dudoso	237	0.8%	
Perdida	107	0.3%	

En la tabla 3, en cuanto a las características financieras de los clientes evaluados y admitidos en la entidad financiera del Perú, el 93.6% del total de clientes se caracterizan como clientes no deudores, durante el periodo de junio 2021 a abril 2022.

También se obtuvo que la antigüedad promedio de los clientes desde que fueron admitidos y evaluados en la entidad financiera es de 5 meses.

Por otro lado, en cuanto a la calificación crediticia de los clientes evaluados en la entidad financiera durante el periodo de estudio, el 96.7% se encontraban con una calificación crediticia normal.

5.1.3. Porcentaje de personas que adquieren el producto.

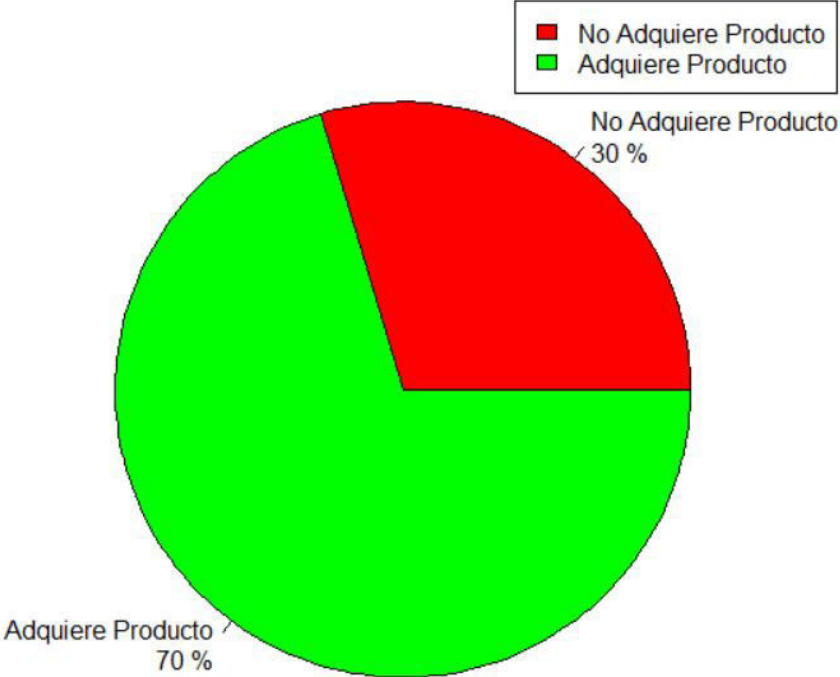


Figura 2. Gráfica de sector circular del porcentaje de clientes según adquisición del producto en una entidad financiera peruana entre el periodo junio 2021 y abril 2022.

En la figura 2, se puede apreciar que, del total de clientes evaluados y admitidos en la entidad financiera del Perú, el 70% adquirieron algún producto durante el periodo de junio 2021 y abril 2022.

5.1.4. Adquisición del Producto Financiero según Características Demográficas

Tabla 4.

Tabla de frecuencia de las características demográficas de los clientes de la entidad financiera del Perú según la adquisición de un producto financiero.

Características	No	%	ME ± D.E	Si	%	ME ± D.E
Sexo						
Masculino	4351	14.1%		11055	35.6%	
Femenino	4845	15.6%		10760	34.7%	
Estado Civil						
Soltero o Sin Pareja	7449	24.0%		20248	65.3%	
Casado	1747	5.6%		1567	5.1%	
Edad			39.15 ± 13.626			34.32 ± 12.180
Rango Monto Ingreso						
Hasta 1025	2052	6.6%		4518	14.6%	
1025 a 2500	4389	14.2%		10221	33.0%	
2500 a 4000	1524	4.9%		3637	11.7%	
4000 a 7000	856	2.8%		2341	7.5%	
7000 a más	375	1.2%		1098	3.5%	
Tipo Ocupación						
Ocup. Dependiente	1097	3.5%		3173	10.2%	
Ocup. Independiente	6754	21.8%		15226	49.1%	
Otros	1345	4.3%		3416	11.1%	
Condición Vivienda						
Alquilada	839	2.7%		2759	8.9%	
De Familiares	3817	12.3%		12540	40.4%	
Propia	4540	14.6%		6516	21.1%	

En la Tabla 4 se observa que, para ambos géneros, masculino y femenino, se obtuvo un mayor porcentaje de clientes masculinos que adquirieron un producto financiero, siendo el 35,6% de los clientes, respecto a las clientas femeninas que adquirieron algún producto financiero.

De acuerdo al estado civil, hay un mayor volumen de clientes solteros o sin pareja y que adquirieron algún producto financiero respecto a los que adquirieron algún producto y se encuentran casados.

También se obtuvo que la edad promedio de los que si adquirieron algún producto financiero es menor a la edad promedio de los que no adquirieron algún producto en la entidad financiera.

En cuanto a la capacidad adquisitiva de monto de ingreso mensual de los clientes se observa que el 33% de los clientes que adquirieron algún producto financiero, se encuentran en el rango de 1025 a 2500. Además, se identificó que el 49% de los clientes evaluados en el estudio, adquirieron algún producto financiero teniendo una ocupación de tipo independiente.

Finalmente, el 40,4% de clientes evaluados que adquirieron algún producto financiero habita en una vivienda de familiares.

5.1.5. Adquisición del Producto Financiero según Características Financieras

Tabla 5.

Tabla de frecuencia de las características financieras de los clientes de la entidad financiera del Perú según la adquisición de un producto financiero.

Características	No	%	ME ± D.E	Si	%	ME ± D.E
Cliente Deudor						
Cliente No Deudor	9185	29.6%		19832	64.0%	
Cliente Deudor	11	0.0%		1983	6.4%	
Antigüedad (meses)			4.90 ± 3.118			5.19 ± 3.063
Calific. Crediticia						
Normal	9185	29.7%		20786	67.0%	
Con problema potencial	5	0.0%		443	1.4%	
Deficiente	2	0.0%		242	0.8%	
Dudoso	3	0.0%		237	0.8%	
Perdida	1	0.0%		107	0.3%	

De la tabla 5 tenemos que el 64% de los clientes evaluados adquirieron algún producto financiero y se caracterizan por no ser deudores.

Además, el número promedio de meses de antigüedad de los clientes que si adquirieron algún producto en la entidad financiera supera al número promedio de meses de antigüedad de los clientes que no adquirieron ningún producto en la entidad financiera durante el periodo de junio 2021 a abril 2022.

5.2. Selección de variables a incluirse en los modelos

Una vez aplicadas las pruebas de chi cuadrado y prueba T de muestras independientes para poder seleccionar las variables que solo deberán considerarse en el modelo logístico, se obtiene lo siguiente:

Tabla 6.

Análisis de variables en relación con la Adquisición de Producto Financiero.

VARIABLE	RELACIÓN	SIG. (P.Valor)	VARIABLE A CONSIDERAR
Sexo	Si influye	0.00	Si
Estado civil	Si influye	0.00	Si
Edad	Si influye	0.00	Si
Rango Monto Ingreso	Si influye	0.00	Si
Tipo Ocupación	Si influye	0.00	Si
Condición Vivienda	Si influye	0.00	Si
Indicador Cliente Deudor	Si influye	0.00	Si
Antigüedad (meses)	Si influye	0.04	Si
Calificación Crediticia	Si influye	0.00	Si

De la tabla 6 podemos decir que las variables si influyen en la adquisición de algún producto financiero en la entidad financiera.

Por tanto, se considerarán las siguientes variables:

- Sexo
- Estado civil
- Edad
- Rango Monto Ingreso
- Tipo Ocupación
- Condición Vivienda
- Indicador Cliente Deudor
- Antigüedad (meses)
- Calificación Crediticia

5.3. Ajuste de Modelo de Regresión Logística

Se utilizará todas las variables ya que resultaron ser influyentes respecto a la variable respuesta “Adquisición del Producto”.

Previo a realizar el modelo logístico se procedió a codificar las variables, tal como se muestra en la tabla 7.

Tabla 7.
Codificaciones de variables categóricas.

Sexo	Masculino	0			
	Femenino	1			
Estado Civil	Soltero o Sin Pareja	0			
	Casado	1			
Rango Monto Ingreso	Hasta 1025	0	0	0	0
	1025 a 2500	1	0	0	0
	2500 a 4000	0	1	0	0
	4000 a 7000	0	0	1	0
	7000 a más	0	0	0	1
Tipo Ocupación	Otros	0	0		
	Ocup. Dependiente	1	0		
	Ocup. Independiente	0	1		
Condición Vivienda	Propia	0	0		
	Alquilada	1	0		
	De Familiares	0	1		
Indicador Cliente Deudor	Deudor	0			
	No Deudor	1			
Calificación Crediticia	Perdida	0	0	0	0
	Con problema potencial	1	0	0	0
	Deficiente	0	1	0	0
	Dudoso	0	0	1	0
	Normal	0	0	0	1

Tabla 8.
Variables en la ecuación del modelo.

	Estimación de β	Error Estándar	Wald	g.l.	Sig.	Exp(β)
Sexo	-0.119	0.026	20.488	1	0.000	0.888
Estado Civil	-0.796	0.041	374.237	1	0.000	0.451
Edad	-0.013	0.001	132.424	1	0.000	0.987
Rango Monto Ingreso			56.915	4	0.000	
1025 a 2500	0.044	0.034	1.695	1	0.193	1.045
2500 a 4000	0.122	0.043	8.022	1	0.005	1.130
4000 a 7000	0.289	0.051	31.992	1	0.000	1.335
7000 a más	0.367	0.070	27.729	1	0.000	1.444
Tipo Ocupación			29.137	2	0.000	
Ocup. Dependiente	0.197	0.050	15.734	1	0.000	1.218
Ocup. Independiente	-0.019	0.038	0.251	1	0.616	0.981
Condición Vivienda			358.741	2	0.000	
Alquilada	0.549	0.047	137.528	1	0.000	1.732
De Familiares	0.553	0.030	333.323	1	0.000	1.738
Indicador Cliente Deudor	-4.196	0.303	191.332	1	0.000	0.015
Antigüedad (meses)	0.011	0.004	6.672	1	0.010	1.011
Calificación Crediticia			103.426	4	0.000	
Con problema potencial	0.053	1.104	0.002	1	0.961	1.055
Deficiente	0.179	1.234	0.021	1	0.885	1.196
Dudoso	-0.337	1.165	0.084	1	0.772	0.714
Normal	-3.130	1.008	9.646	1	0.002	0.044
Constante	8.215	1.054	60.698	1	0.000	3694.924

De la tabla 8 tenemos lo siguiente:

Existe 1.045 mayor probabilidad que el cliente adquiera un producto financiero cuando su ingreso mensual se encuentra en el rango de 1025 a 2500 soles, respecto a cuando su ingreso mensual se encuentra en el rango menor a 1025 soles.

A su vez, se obtuvo que es 1.218 más probable que el cliente adquiriera un producto en la financiera cuando su tipo de ocupación es dependiente respecto a algún otro tipo de ocupación eventual.

También, se identificó que es 1.732 más probable que el cliente adquiriera un producto financiero cuando su tipo de vivienda es alquilada respecto a cuando reside en una vivienda propia.

Además, es 1.055 más probable que el cliente adquiriera un producto en la financiera si su calificación crediticia se encuentra con problemas potenciales respecto a cuando su calificación crediticia está en pérdida.

Por otro lado, se observa que las variables si resultan ser significativas, siendo estas:

Sexo, estado civil, edad, rango del monto ingreso, tipo de ocupación, condición de vivienda, indicador de cliente deudor, antigüedad del cliente y la calificación crediticia del cliente.

Por ello, considerando las variables mencionadas las cuales serán seleccionadas de forma definitiva según los resultados de la tabla 8, la ecuación de la regresión logística para la propensión de adquisición de un producto financiero en la entidad financiera es la siguiente:

$$P(\text{Adquirir Producto} = Si) = \frac{1}{1 + e^{-R}}$$

donde:

$$\begin{aligned} R = & \beta_0 + \beta_1 \times \text{Sexo} + \beta_2 \times \text{Estado Civil} + \beta_3 \times \text{Edad} + \beta_4 \times \text{Rango Monto Ingreso} \\ & + \beta_5 \times \text{Tipo Ocupación} + \beta_6 \times \text{Condición de Vivienda} \\ & + \beta_7 \times \text{Indicador Cliente Deudor} + \beta_8 \times \text{Antigüedad (meses)} \\ & + \beta_9 \times \text{Calificación Crediticia} \end{aligned}$$

5.4. Bondad de ajuste del modelo: Propensión de adquisición de un producto financiero

Significancia del modelo.

Para evaluar la significancia de los coeficientes del modelo se aplica la siguiente prueba de Hipótesis:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_{10} = 0$$

$$H_1: \text{Por lo menos un } \beta_i \neq 0, \forall i = 1, 2, 3, \dots, 10$$

Tabla 9.
Prueba sobre los coeficientes del modelo.

Valor chi-cuadrado	g.l.	Sig.
52,714	8	0,000

De la tabla 9, se concluye que con 5% de significancia, el modelo es significativo.

5.5. Matriz de Confusión

Tabla 10.
Tabla de clasificación del modelo de regresión logística para la propensión de adquisición de un producto financiero en la entidad financiera.

Observado	Pronosticado		Porcentaje correcto	
	Adquiere Producto No	Adquiere Producto Si		
Adquiere Producto	No	1256	7940	13.7%
	Si	1017	20798	95.3%
Porcentaje global			71.1%	

De acuerdo a la tabla 10, el modelo clasifica de forma correcta el 71.1% de los casos de que un cliente adquiera un producto financiero dentro de la entidad financiera peruana.

VI. CONCLUSIONES

Según los análisis realizados, se logró determinar que las características consideradas en el trabajo si se encuentran relacionadas con la adquisición de algún producto financiero dentro de la entidad financiera peruana, principalmente de acuerdo a las características demográficas y financieras de los clientes incluidas en el estudio.

Se determinó que para que un cliente adquiriera un producto financiero influye su rango de monto de ingreso mensual, su tipo de ocupación, su condición de vivienda, así como la calificación crediticia.

Se generó un modelo logístico el cual con un 5% de significancia se pudo determinar que el modelo es significativo, por lo que el modelo diseñado permite cumplir el objetivo de medir la propensión de adquisición de algún producto financiero de la cartera de clientes en la entidad financiera peruana a través de las características demográficas y financieras que estos presentan.

Los resultados obtenidos en el análisis realizado son alentadores ya que se obtuvo un modelo empleado capaz de alcanzar una buena precisión en la propensión a la adquisición de algún producto financiero.

VII. RECOMENDACIONES

Se sugiere en el desarrollo del modelo, que este puede seguir mejorando en la medida que se consideren e incorporen más variables de las características financieras de los clientes, como por ejemplo información que tienen los clientes en otras entidades.

Con esto, se recomienda generar y comparar más modelos logísticos que incluyan más características de los clientes y así seguir en la búsqueda de enriquecer la medición de propensión de adquisición de algún producto financiero por parte de los clientes de la entidad financiera.

Para trabajos posteriores también se recomienda emplear una gran cantidad histórica de datos de clientes para mantener la consistencia del modelo logístico predictivo.

Se recomienda seguir en búsqueda de nuevos modelos que mejoren la precisión y predicción en cuanto a la adquisición de productos financieros, considerando también otros agentes externos como campañas o estrategias.

VIII. BIBLIOGRAFÍA

- Ayala, J. (2020). *Modelo de Propensión a la toma de Créditos de Consumo en una Empresa del Sector Financiero*. (Tesis de Maestría). Universidad del Desarrollo. Santiago de Chile, Chile.
- Barrueta, R., y Castillo, E. (2018). *Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos*. (Tesis de Pregrado). UPC. Lima, Perú.
- Camarero, A., Almazán, A. & Mañas, B. *Regresión Logística. Fundamentos y aplicación a la investigación sociológica*. (Investigación). UNED. España.
- Chipana, Y. (2018). *Evaluación de la gestión crediticia en el otorgamiento de créditos en la Financiera Credinka Agencia Yunguyo. Periodo 2015 – 2016*. (Tesis de Pregrado). Universidad Nacional del Altiplano. Puno, Perú.
- Esguerra, A., y Niño, D. (2017). *Propensión de los clientes de tarjeta de crédito del banco de Bogotá hacia la educación financiera*. (Trabajo de Investigación). Uniminuto. Bogotá, Colombia.
- Garson, D. (2014). *Logistic Regression: Binary & Multinomial*.
- Gross, A. (2018). *Elaboración de un modelo de propensión al crédito de consumo capaz de discriminar a nivel de individuo utilizando solamente información financiera de carácter pública*. (Tesis de Maestría). Pontificia Universidad Católica de Chile. Santiago de Chile, Chile.
- Hosmer, D. & Lemeshow, S. (2013). *Applied Logistic Regression*.
- Zarabia, C. (2020). *Identificación de la propensión a la adquisición de un subproducto de una tarjeta de crédito en una entidad bancaria*. (TSP de Pregrado). UNALM. Lima, Perú.

IX. ANEXOS

Anexo 1: Código

```
library(tidyverse)
library(PerformanceAnalytics)
library(corrplot)
library(psych)
library(GGally)
library(gridExtra)
library(lmtest)
library(car)

data<-read.csv("datatesisR.csv",header=TRUE,sep=";",dec=".")
head(data)
str(data)
datos <- data[,c(1:10)]
head(datos)
table(data$FlagAdquiereProducto)
### Resumen
summary(data)
## describe
library(Hmisc)
describe(data)
library(graphics)
porcentajes<-round(100*table(datos$FlagAdquiereProducto)/31011,digits = 0)
porcentajes<-paste(round(100*table(datos$FlagAdquiereProducto)/31011,digits = 0),rep("%",2))
etiquetas<-paste(c("No Adquiere Producto", "Adquiere Producto"),porcentajes,sep = "\n")
pie(table(data$FlagAdquiereProducto),labels = etiquetas, col = c("red", "green"))
x11()
legend(x = "topright", # Posición
       legend = c("No Adquiere Producto", "Adquiere Producto"), # Leyenda
       fill = c("red", "green"), border = "black") # Color de leyenda
```

```
modelo<-lm(data = datos,datos$FlagAdquiereProducto~.)
summary(modelo)
```

```
vif(modelo)
```

```
> vif(modelo)
```

	GVIF	Df	GVIF^(1/(2*Df))
Sexo	1.017671	1	1.008797
EstadoCivil	1.131435	1	1.063690
Antiguedad	1.033716	1	1.016718
CondicionVivienda	1.245506	2	1.056420
Ind_Cli_Deu	1.074244	1	1.036457
Edad	1.334369	1	1.155149
CalifiCrediticia	1.073823	4	1.008943
EstadoOcupacion	1.059085	2	1.014455
RangoMontoIngreso	1.061454	4	1.007483

```
# Errores
```

```
qqnorm(modelo$residuals)
```

```
qqline(modelo$residuals)
```

Anexo 2. R-squared del modelo

Residual standard error: 0.4356 on 30993 degrees of freedom

Multiple R-squared: 0.0909, Adjusted R-squared: 0.0904

F-statistic: 182.3 on 17 and 30993 DF, p-value: < 2.2e-16