

Gestión del riesgo crediticio: comparación de modelos logit y probit bajo un enfoque clásico y bayesiano

Credit risk management: comparison of logit and probit models under classical and bayesian approaches

Richard Fernando Fernández Vásquez¹

RESUMEN

En el ámbito de la gestión del riesgo crediticio, el credit scoring constituye una herramienta estadística fundamental que permite a las instituciones financieras mejorar sus decisiones de aprobación o rechazo de créditos, ajustando sus políticas al perfil de riesgo de cada solicitante. En respuesta a este contexto, el objetivo general de la presente investigación fue comparar los modelos estadísticos de regresión logit y probit bajo los enfoques clásico y bayesiano, con la finalidad de desarrollar un modelo de credit scoring aplicado a solicitantes de préstamos personales, orientado a segmentarlos y proponer acciones según su nivel de riesgo. La investigación tuvo un enfoque cuantitativo; el método empleado fue hipotético-deductivo, con un diseño no experimental, transversal, descriptivo y correlacional. La población estuvo conformada por 5 584 solicitantes de préstamos personales, tanto cumplidores como incumplidores, cuyos datos fueron obtenidos de la plataforma DataCamp como parte de una iniciativa académica orientada al análisis del riesgo crediticio. Para el desarrollo del estudio, se establecieron una muestra de entrenamiento y una de validación, correspondientes al 70 % y 30 % de los datos, respectivamente. Se compararon los modelos estadísticos logit y probit estimados bajo los enfoques clásico y bayesiano. Los resultados evidenciaron que, al considerar los indicadores de desempeño —precisión, F1-score, área bajo la curva ROC e índice de Gini—, el modelo de regresión logit bajo el enfoque bayesiano presentó el mejor

Recibido: 08/07/2025
Aceptado: 31/01/2026
Publicado: 31/03/2026

¹Universidad Nacional
de Ingeniería, Lima, Perú

Correspondencia:
rfernandezv@uni.edu.pe
<https://orcid.org/0000-0003-1721-8527>

Licencia:



Revista de la Facultad de
Ingeniería Económica,
Estadística y Ciencias
Sociales de la Universidad
Nacional de Ingeniería

rendimiento, con valores de 0.6201, 0.6310, 65.6286 y 31.2573, respectivamente.

Palabras clave: incumplimiento de crédito, credit scoring, regresión logit, regresión probit, enfoque clásico, enfoque bayesiano, MCMC.

ABSTRACT

In the field of credit risk management, credit scoring is a fundamental statistical tool that allows financial institutions to improve their decisions on whether to approve or reject loans, adjusting their policies to the risk profile of each applicant. In response to this context, the overall objective of this research was to compare the logit and probit statistical regression models under the classical and Bayesian approaches, with the aim of developing a credit scoring model applied to personal loan applicants, aimed at segmenting them and proposing actions according to their level of risk. The research took a quantitative approach; the method used was hypothetical-deductive, with a non-experimental, cross-sectional, descriptive, and correlational design. The population consisted of 5,584 personal loan applicants, both compliant and non-compliant, whose data were obtained from the DataCamp platform as part of an academic initiative aimed at credit risk analysis. For the development of the study, a training sample and a validation sample were established, corresponding to 70% and 30% of the data, respectively. The logit and probit statistical models estimated under the classical and Bayesian approaches were compared. The results showed that, when considering the performance indicators—accuracy, F1-score, area under the ROC curve, and Gini index—the logit regression model under the Bayesian approach performed best, with values of 0.6201, 0.6310, 65.6286, and 31.2573, respectively.

Keywords: credit default, credit scoring, logit regression, probit regression, classical approach, bayesian approach, MCMC.

1. INTRODUCCIÓN

1.1 Justificación e importancia

El fundamento teórico de este estudio se sustenta en la implementación de fundamentos científicos estadísticos para comparar dos enfoques inferenciales distintos —clásico y bayesiano— mediante los modelos de regresión logit y probit, con el propósito de fortalecer la formulación de un

modelo de predicción de riesgo crediticio como herramienta clave para su gestión. La selección del modelo de regresión con mayor capacidad de discriminación permite cuantificar con precisión la tasa de ocurrencia de default de los solicitantes de préstamos personales, lo que contribuye a la construcción de un modelo de credit scoring más robusto y confiable. Asimismo, este enfoque facilita la segmentación de los solicitantes según su nivel de riesgo y la formulación de acciones diferenciadas, orientadas a una toma de decisiones crediticias más eficiente y sustentada en evidencia estadística.

La justificación práctica de esta investigación resulta relevante en la esfera de la administración de riesgos crediticios, ya que proporciona una herramienta estadística que otorga a las instituciones bancarias optimizar sus decisiones de aprobación o rechazo de créditos, ajustando sus políticas al nivel de probabilidad de incumplimiento para cada cliente potencial. Este enfoque se enmarca de acuerdo con lo estipulado por la normativa para la Gestión del Riesgo de Crédito, ratificado por la Resolución SBS N.° 3780-2011 y sus normas de carácter modificatorio, al igual que en la implementación de la normativa para la Gestión de Riesgos de Modelo, ratificado mediante la normativa SBS N.° 00053-2023.

Asimismo, la investigación aporta una herramienta concreta que facilita a las entidades financieras la segmentación de sus clientes bajo un enfoque estadístico, permitiendo la aplicación de acciones diferenciadas de acuerdo con el nivel de riesgo identificado. De igual manera, la Superintendencia de Banca, Seguros y AFP (SBS) se verá beneficiada con los resultados de este estudio, en la medida en que contribuirán a un monitoreo más preciso de la posibilidad de incumplimiento de crédito, elaboración de reglamentos con mayor impacto y a la promoción de prácticas crediticias responsables.

La Asociación de Bancos del Perú (ASBANC) se verá beneficiada, dado que podrá emplear los resultados de esta investigación como un caso de uso para el despliegue de un modelo de riesgo crediticio dentro de los procesos de innovación digital del sector bancario, contribuyendo a la mejora de sus estudios técnicos y al fortalecimiento de sus políticas de autorregulación.

Finalmente, la presente investigación beneficiará a investigadores de distintas universidades del Perú y del extranjero, al proporcionar un estudio comparativo sobre la aplicación de herramientas estadísticas empleadas

para la elaboración de sistemas de credit scoring, enriqueciendo el conocimiento académico en el contexto del control ante la incertidumbre crediticia.

El soporte metodológico del estudio se sustenta en lo señalado por Dassatti (2019), quien indicó que los modelos de regresión logit y probit son estadísticamente adecuados para el desarrollo de sistemas de credit scoring, debido a su efectividad en la evaluación de probabilísticos ante el default. De manera complementaria, Banca & Economía (2022) destaca que, entre los modelos estadísticos más utilizados y relevantes para la construcción de credit scoring, sobresalen la regresión logit y la regresión probit. Adicionalmente, la investigación incorpora dos enfoques inferenciales —clásico y bayesiano—, los cuales son comparados a través de métricas de desempeño como la sensibilidad, la precisión, el F1-score, AUC y el índice de Gini. En función de las justificaciones expuestas, la trascendencia del estudio desarrollado que enfoca en el objetivo de estimar, con el mayor grado de precisión posible, la probabilidad de incumplimiento de los solicitantes de préstamos personales, aplicando principios de la ciencia estadística mediante la revisión comparativa de los modelos de regresión logit y probit bajo los enfoques clásico y bayesiano, contribuyendo así al fortalecimiento del manejo probabilístico ante el default en entidades del sistema financiero.

1.2 Antecedentes

Reyes y Sosa (2022) indicaron que, para las instituciones financieras, el riesgo a la exposición al incumplimiento crediticio es uno de los temas más relevantes con los que el sistema financiero en su conjunto (parte de supervisión y regulación, como la SBS en el Perú) se debe preocupar. En ese orden de ideas, los autores plantearon un modelo de credit scoring. Es así que, emplearon una base de datos de clientes que contaban con el producto de tarjeta de crédito y aplicaron la regresión logística para predecir el riesgo de incumplimiento del solicitante. Por último, recomiendan que este tipo de trabajos se hagan en todas las entidades financieras, dado lo que aporta a prevenir el riesgo de pérdida crediticia a nivel macroeconómico, lo que finalmente beneficia al sistema financiero en su conjunto. Según Banca & Economía (2022), el credit scoring es un conjunto de técnicas estadísticas con las cuales se puede estimar la tasa de mora de un portafolio crediticio. Para el cálculo de dicha tasa se emplea una variedad de modelos estadísticos, entre los que destacan la regresión logit y

la regresión probit. Por otro lado, en el ámbito de la gestión de riesgos de créditos, se indica que esta labor debe ser constante, y el credit scoring un arma fundamental en la concesión de préstamos personales. Su uso ayuda a reducir la exposición a pérdidas financieras originadas en los impagos de los clientes.

Dassatti (2019) explica que los modelos estadísticos utilizados en el desarrollo del credit scoring son la base de la mayoría de las decisiones que toman las instituciones financieras del sector basadas en el otorgamiento de préstamos personales. Además, señala, que las técnicas más empleadas son los modelos de regresión logit y probit, por su eficiencia en este tipo de aplicaciones.

Hilbck (2022) comentó que la evaluación crediticia de los solicitantes de préstamos, a través de modelos estadísticos aplicados al desarrollo del credit scoring, permite estimar la probabilidad de que un sujeto no cumpla con sus obligaciones crediticias. Esto hace que las entidades financieras puedan tomar decisiones más acertadas, lo que a su vez, contribuye a optimizar la gestión del riesgo crediticio.

Webster (2011) utilizó el modelo estadístico de regresión logística bayesiana con el objetivo de elaborar modelos de credit scoring, resaltando la importancia de la incorporación de distribuciones a priori. Estas distribuciones permiten aumentar la precisión en la identificación de prestatarios con alta probabilidad de impago de sus obligaciones crediticias. Además, el autor sostiene que, cuando se dispone de grandes bases de datos, los resultados obtenidos con la regresión logística bajo el enfoque bayesiano se asemejan a los del enfoque clásico, sin diferencias importantes en cuanto a la precisión.

Nopper (2020) afirmaron que los modelos de credit scoring son una herramienta eficaz para la reducción de los costos operativos asociados al otorgamiento de préstamos bajo la modalidad de microcrédito. En su trabajo, analizaron una base de datos de 469 996 clientes, utilizando variables tanto cuantitativas como cualitativas, y evaluaron el desempeño predictivo del modelo de regresión logística. Los resultados, medidos a través del índice de Gini, demostraron que dicho modelo identifica de manera efectiva a los solicitantes con mayor ratio de mora de sus obligaciones crediticias. Agrawal et al. (2021) evaluaron la calificación crediticia de los clientes mediante modelos de aprendizaje automático. Los resultados evidenciaron que el modelo estadístico de regresión logística presentó la mayor precisión en la identificación de solicitantes con alta probabilidad de incumplimiento.

miento crediticio, en comparación con otros modelos de machine learning. Finalmente, Mousavi y Gazori (2023) realizaron una revisión de modelos de calificación crediticia, en la cual destacan el uso del modelo estadístico de regresión logística como una de las técnicas más relevantes para el desarrollo de sistemas de credit scoring.

1.3 Objetivos

Objetivo General

Comparar los modelos estadísticos de regresión logit y probit, bajo los enfoques clásico y bayesiano, con el objetivo de elaborar un sistema de evaluación crediticia aplicado a solicitantes de préstamos personales, con el fin de segmentarlos y proponer acciones diferenciadas según su nivel de riesgo.

Objetivos Específicos

Analizar la eficacia de los modelos estadísticos de regresión logit, bajo los enfoques clásico y bayesiano, para la estimación de la ratio de mora de los solicitantes de préstamos personales, utilizando como indicadores la sensibilidad, la precisión, el F1-score, el área bajo la curva ROC y el índice de Gini.

Analizar la eficacia de los modelos estadísticos de regresión probit, bajo los enfoques clásico y bayesiano, para la estimación de la ratio de mora de los solicitantes de préstamos personales, utilizando como indicadores la sensibilidad, la precisión, el F1-score, el área bajo la curva ROC y el índice de Gini.

1.4 Hipótesis Central de Investigación

Hipótesis nula (H_0)

El modelo estadístico de regresión logit bajo el enfoque bayesiano no presenta un desempeño predictivo superior en la creación de un sistema de evaluación crediticia, ya que no evidencia valores significativamente mayores en los indicadores de sensibilidad, precisión, F1-score, área bajo la curva ROC y coeficiente de Gini, lo que limita su capacidad para lograr una mejor segmentación del riesgo crediticio.

Hipótesis alternativa (H₁)

El modelo estadístico de Regresión Logit bajo el enfoque bayesiano presenta un desempeño predictivo superior en la creación de un sistema de evaluación crediticia, reflejado en valores significativamente mayores de sensibilidad, precisión, F1-score, área bajo la curva ROC y coeficiente GINI, permitiendo una mejor segmentación del riesgo crediticio.

2. METODOLOGÍA

2.1 ENFOQUE DE INVESTIGACIÓN

La orientación del estudio fue de tipo cuantitativo, ya que se empleó información sociodemográfica y de comportamiento crediticio de los solicitantes de préstamos personales. A partir de estos datos, se realizó un análisis estadístico mediante los modelos de regresión logit y probit, estimados bajo los enfoques clásico y bayesiano. Finalmente, se evaluó su desempeño a través de la comparación de indicadores estadísticos de rendimiento.

2.2 MÉTODO DE INVESTIGACIÓN

El enfoque metodológico implementado en esta investigación fue el hipotético-deductivo, ya que la investigación se fundamentó en una hipótesis central, a partir de la cual se desarrolló el proceso investigativo, el mismo que posteriormente fue sometido a un riguroso proceso de validación.

2.3 DISEÑO DE INVESTIGACIÓN

El diseño del actual trabajo fue:

- No experimental, ya que no se realizó ninguna intervención directa sobre los solicitantes de préstamos personales; las variables no fueron manipuladas y el análisis se limitó al procesamiento de la información posterior a su descarga. Asimismo, el estudio fue de tipo retrospectivo, dado que los datos de los solicitantes que incumplieron y no incumplieron el crédito fueron establecidos con anterioridad, sin pretender determinar una conexión de causalidad entre el resultado y los factores explicativos.
- Transversal, puesto que se utilizó información correspondiente a los solicitantes de préstamos personales en un momento específico dentro de un determinado periodo de tiempo.
- Descriptivo, debido a que la investigación se centró en el análisis estadístico de los solicitantes de préstamos personales que incumplieron y no

incumplieron el crédito, mediante el uso de los modelos estadísticos de regresión logit y probit, estimados bajo los enfoques clásico y bayesiano.

- Correlacional, ya que se buscó establecer una relación matemática-estadística entre las variables asociadas a los solicitantes de préstamos personales y el incumplimiento del crédito.

2.4 POBLACIÓN Y MUESTRA

La población objeto de la presente investigación estuvo conformada por los solicitantes de préstamos personales de una entidad financiera. La muestra estuvo constituida por 5,584 registros correspondientes a solicitantes que incumplieron y no incumplieron con el pago de sus créditos. Dichos datos fueron recopilados de la plataforma DataCamp, disponibilizados por Moretz, como parte de una iniciativa académica orientada al análisis del riesgo crediticio.

Para el desarrollo de la investigación se trabajó con la totalidad de la muestra, la cual fue dividida de la siguiente manera:

- Muestra de entrenamiento: conformada por 3,918 registros (70% del total), utilizada para la estimación de los modelos estadísticos de regresión logit y probit bajo los enfoques clásico y bayesiano.
- Muestra de validación: integrada por 1,666 registros (30% del total), empleada para evaluar el desempeño predictivo de los modelos mediante los indicadores de sensibilidad, precisión, F1-score, área bajo la curva ROC e índice de Gini.

2.5 VARIABLES

La variable dependiente fue:

- `loan_status`: variable dicotómica que toma el valor de 1 si el solicitante incumplió con el pago del préstamo personal y 0 en caso contrario.

Las variables independientes fueron:

- `loan_amnt`: es el monto del préstamo personal requerido por el solicitante.
- `int_rate`: es la tasa de interés que ha generado el solicitante.
- `grade`: el grado es el puntaje asignado al solicitante, donde A: indica la clase más alta de solvencia y G: la más baja. Esta puntuación de la oficina de la entidad financiera, refleja el historial de crédito del individuo y es la única variable de comportamiento.
- `emp_length`: es la duración de la ocupación del peticionario.

- home_ownership: es el estado de la pertenencia de vivienda que tiene el peticionario.
- annual_inc: es el ingreso anual del solicitante.
- age: es la edad del solicitante en años.

2.6 MODELO ESTADÍSTICO DE REGRESIÓN LOGIT

2.6.1 ENFOQUE CLÁSICO

Según Wooldridge (2010), el modelo logit es una regresión usada cuando la variable dependiente Y es binaria. Transforma una regresión lineal en una estimación de probabilidades de una variable dependiente Y binaria.

Matemáticamente, las probabilidades haciendo uso del modelo logit quedan expresadas de la siguiente manera:

$$Prob(Y = 1|X) = f(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) = f(z)$$

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

Donde:

- $Prob(Y=1|X)$: corresponde a la probabilidad asociada a la ocurrencia del evento de interés dado un conjunto de variables
- $\beta_0, \beta_1, \beta_2, \dots, \beta_k$: son los coeficientes del modelo, que se valoran mediante el procedimiento de máxima verosimilitud.
- $f(z)$: es la función logística de distribución acumulada

Finalmente, el modelo logit queda expresado de la siguiente manera:

$$\log\left(\frac{Prob(Y = 1|X)}{1 - Prob(Y = 1|X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Donde:

- $\frac{Prob(Y = 1|X)}{1 - Prob(Y = 1|X)}$: es el odds ratio

La evaluación de los coeficientes se realiza mediante el Estimador de Máxima Verosimilitud.

2.6.2 ENFOQUE BAYESIANO

Gelman et al. (2020), señalan que el modelo estadístico de regresión logit bajo el enfoque bayesiano permite mejorar la estimación de los coeficientes del modelo al incorporar información previa o a priori y aplicar el teorema de Bayes. A diferencia del enfoque clásico, donde los coeficientes se consideran constantes, bajo el enfoque bayesiano los coeficientes son considerados como variables aleatorias, asignándole una distribución de probabilidades a priori. Posteriormente, los coeficientes se estiman a partir de su distribución a posterior, la cual resulta de combinar la distribución a priori con los datos observados. Esta estimación queda representada mediante la expresión matemática:

$$Prob(\beta | \text{datos}) = \frac{Prob(\text{datos} | \beta)P(\beta)}{P(\text{datos})}$$

Donde:

- $P(\beta)$: es la distribución a priori.
- $Prob(\text{datos} | \beta)$: es la verosimilitud.
- $Prob(\beta | \text{datos})$: es la distribución a posteriori.

Lo señalado por Gelman et al. (2020) también se emplea para el modelo estadístico de regresión probit bajo el enfoque bayesiano.

Gelman et al. (2013) explican que a fin de la estimación de los coeficientes del modelo estadístico de regresión logit bajo el enfoque bayesiano se realiza mediante simulación, utilizando el procedimiento de Monte Carlo vía Cadenas de Markov (MCMC), concretamente por medio del algoritmo de Metropolis-Hastings. En este enfoque, se ocupa una distribución normal a priori hacia los coeficientes β , expresada como:

$$\beta \sim N(\mu_0, \Sigma_0)$$

El procedimiento MCMC con el algoritmo de Metropolis-Hastings permite generar muestras de la distribución posterior de los coeficientes y sigue la siguiente secuencia de pasos:

- a. Inicializar los coeficientes $\beta^{(0)}$
- b. Iterar generando una cadena de valores $\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(T)}$, de modo que:

$$\beta^{(t)} \sim P(\beta | \text{datos})$$

- c. Analizar la cadena para obtener la media posterior de cada coeficiente, es decir $E(\beta_j | \text{datos})$.

Martin et al. (2011), señalan que la ejecución del modelo estadístico de regresión logit bajo el enfoque bayesiano usando MCMC mediante el algoritmo de Metropolis-Hastings, se encuentra disponible a través de la función MCMClogit, perteneciente al paquete MCMCpack en el software estadístico R.

2.7 MODELO ESTADÍSTICO DE REGRESIÓN PROBIT

2.7.1 ENFOQUE CLÁSICO

Según Wooldridge (2010), el probit es una regresión usada también cuando la variable dependiente Y es binaria. Transforma una regresión lineal en una estimación de probabilidades de una variable dependiente Y binaria.

Matemáticamente, las probabilidades haciendo uso del modelo probit quedan expresadas de la siguiente manera:

$$\text{Prob}(Y = 1|X) = f(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) = f(z)$$

$$f(z) = \int_{-\infty}^z \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{u^2}{2}} du$$

Donde:

- $\text{Prob}(Y=1|X)$: es la probabilidad de que acontezca el suceso de interés dado un conjunto de variables
- $\beta_0, \beta_1, \beta_2, \dots, \beta_k$: son los coeficientes del modelo, que se valoran mediante el método de máxima verosimilitud.
- $f(z)$: es la función normal estándar de distribución acumulada

Finalmente, el modelo probit queda expresado de la siguiente manera:

$$F(\text{Prob}(Y=1|X))^{-1} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Donde:

- $F(\text{Prob}(Y=1|X))^{-1}$: es la función inversa de la función normal estándar de distribución acumulada

La evaluación de los coeficientes se realiza mediante el Estimador de Máxima Verosimilitud.

2.7.2 ENFOQUE BAYESIANO

Albert y Chib (1993) explican que a fin de la valoración de los coeficientes del modelo estadístico de regresión probit bajo el enfoque bayesiano se realiza mediante simulación, utilizando el Método de Monte Carlo vía Cadenas de Markov (MCMC), específicamente mediante el muestreo de Gibbs. En este enfoque, ocupa una distribución normal a priori hacia los coeficientes β , expresada como:

$$\beta \sim N(\mu_0, \Sigma_0)$$

El procedimiento MCMC mediante el muestreo de Gibbs permite generar muestras de la distribución posterior de los coeficientes y sigue la siguiente secuencia de pasos:

- a. Inicializar introduciendo una variable latente z_i , de tal manera que:

$$z_i | \beta, y_i \sim \begin{cases} N(x'_i \beta, 1) \text{ truncada inferior en } 0, & \text{si } y_i = 1 \\ N(x'_i \beta, 1) \text{ truncada superior en } 0, & \text{si } y_i = 0 \end{cases}$$

- b. Actualizar los coeficientes β a partir de una distribución posterior:

$$\beta | \text{datos} \sim N(\mu_\beta, \Sigma_\beta)$$

- c. Repetir los pasos a y b para generar una Cadena de Markov convergente a la distribución posterior de β .
- d. Analizar la cadena para obtener la media posterior de cada coeficiente, es decir $E(\beta_j | \text{datos})$.

Martin et al. (2011), señalan que la consumación del modelo estadístico de regresión probit bajo el enfoque bayesiano usando MCMC mediante el muestreo de Gibbs, se encuentra disponible a través de la función MCMCprobit, perteneciente al paquete MCMCpack en el software estadístico R.

2.8 INDICADORES DE PREDICCIÓN

2.8.1 MATRIZ DE CONFUSIÓN

Barrios (2019), señala que la matriz de confusión es un instrumento útil a fin de valorar la actuación de un modelo de regresión cuya variable dependiente es dicotómica. En dicha matriz, las columnas encarnan el número de registros en la clase pronosticada, en tanto que las filas encarnan el número de registros en la clase observada.

Tabla 1
Matriz de confusión

Observado	Pronosticado			Porcentaje correcto
	Evento		Porcentaje correcto	
	0	1		
Evento	0	a	b	$a/(a+b)$
	1	c	d	$d/(c+d)$
Exactitud				$(a+d)/(a+b+c+d)$

A continuación, se ostenta la representación de cada término de la matriz de confusión:

- a: es el dígito de pronósticos correctos de clase 0, toma el alias de verdaderos negativos (VN).
- b: es el dígito de pronósticos incorrectas de clase 1, toma el alias de falsos positivos (FP).
- c: es el dígito de pronósticos incorrectas de clase 0, toma el alias de falsos negativos (FN).
- d: es el dígito de pronósticos correctas de clase 1, toma el nombre de verdaderos positivos (VP).

Asimismo, se presenta los indicadores estadísticos asociados a la matriz de confusión:

- Sensibilidad = $d/(c+d)$, evidencia la potestad que conserva un modelo a fin de constituir convenientemente la categoría de interés en la clase observada de la variable dependiente.
- Especificidad = $a/(a+b)$, revela la extensión que conserva un modelo a fin de constituir convenientemente la categoría que no es de utilidad de la variable dependiente.
- Precisión = $d/(b+d)$, revela la potestad que posee un modelo a fin de constituir convenientemente la categoría que es de interés en la clase pronosticada de la variable dependiente.
- Exactitud = $(a+d)/(a+b+c+d)$, muestra la potestad que posee un modelo a fin de constituir convenientemente de forma integral la variable dependiente.
- F1-score = $2 \times \text{Precisión} \times \text{Sensibilidad} / (\text{Precisión} + \text{Sensibilidad})$, es una métrica que resume la precisión y sensibilidad.

Finalmente, el modelo estadístico que muestre mayores valores de sensibilidad, precisión y F1-score es el modelo más conveniente para el pronóstico de la variable dependiente.

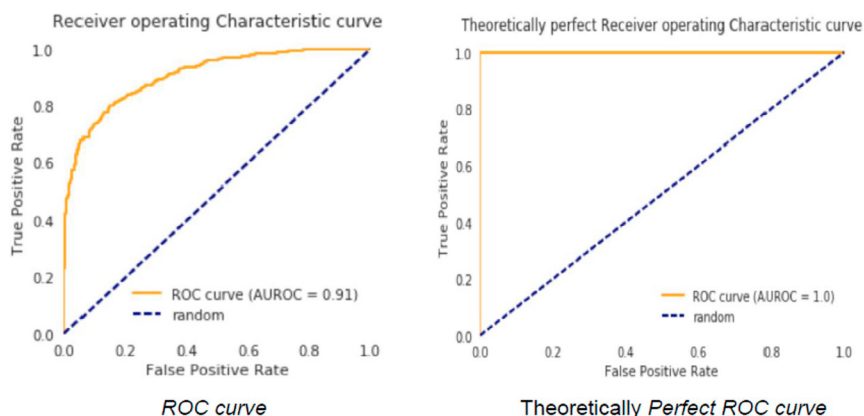
2.8.2 CURVA ROC

Macskassy y Foster (2003) señalan que la curva ROC (Receiver Operating Characteristic) es una herramienta ampliamente utilizada para evaluar el desempeño de modelos cuya variable dependiente es dicotómica. En este contexto, Adams (1999) explica que la curva ROC se construye representando, en el eje vertical, la proporción de observaciones correctamente clasificadas como clase 0 (tasa de verdaderos negativos) y, en el eje horizontal, la proporción de observaciones incorrectamente clasificadas como clase 1 (tasa de falsos positivos).

Una curva ROC ideal clasifica correctamente todos los casos de la clase 1 como clase 1, lo que implica una sensibilidad y una especificidad iguales a 1. Por lo tanto, cuanto más cercano sea el valor del área bajo la curva ROC a 1, mejor será el desempeño predictivo del modelo.

En la Figura 1 se muestra un área bajo la curva ROC de 0.92 correspondiente a un modelo específico, comparándose con el valor de 1 de una curva ROC ideal, representativa de un modelo con capacidad predictiva perfecta.

Figura 1
Curva ROC



Nota. Tomado de Calibration of Machine Learning Classifiers for Probability of Default Modelling, por P. Fonseca y H. Lopes, 2017, James Finance.

2.8.3 INDICE DE GINI

Adams y Hand (1999), así como Hand y Anagnostopoulos (2013), mencionan que el índice de Gini también se utiliza para evaluar el desempeño de modelos cuya variable dependiente es dicotómica, constituyendo una alternativa al uso del área bajo la curva ROC. En este sentido, cuanto más próximo sea su valor a 1, mejor será el desempeño predictivo del modelo. La fórmula matemática para su cálculo es la siguiente:

$$\text{GINI}=2(\text{ROC} - 0.5)$$

2.9 CONSTRUCCIÓN DEL CREDIT SCORE

Dassatti (2019) señala que, una vez identificado el modelo estadístico más adecuado para estimar la probabilidad de incumplimiento, dichas probabilidades pueden transformarse en puntuaciones crediticias, conocidas como credit score, mediante un proceso de reescalamiento. En este contexto, Thomas et al. (2002) proponen la siguiente fórmula matemática para calcular el credit score a partir de la probabilidad de incumplimiento:

$$\text{Credit Score} = \text{Score}_0 + \text{PDO} \frac{\log_2(\text{odds})}{\log_2(\text{odds}_0)}$$

Donde:

- Score_0 : es el score base para duplicar el odds, por lo general toma el valor de 600.
- PDO : son los puntos para duplicar el odds, por lo general toma el valor de 20.
- Odds : se calcula como $(1-\text{PD})/\text{PD}$, siendo PD, la probabilidad de inobservancia a fin de cada registro de la base de datos.
- Odds_0 : es el odds de referencia, por lo general toma el valor de 50.

Espin y Rodríguez (2013) señalan que, una vez obtenidas las puntuaciones crediticias (credit score) a partir de las probabilidades de incumplimiento, estas pueden agruparse en dos o más categorías con el objetivo de segmentar a los solicitantes según su nivel de riesgo crediticio.

2.10 PROCEDIMIENTO ESTADÍSTICO

Las actividades correspondientes al procedimiento estadístico aplicado en la presente investigación, que dieron lugar a los hallazgos, fueron las

siguientes:

- En primera instancia, se segmentó la base de datos completa de clientes del banco en un conjunto de entrenamiento y otro de validación, con proporciones de 70 % y 30 %, respectivamente.
- En segunda instancia, con la muestra de entrenamiento, se aplicaron los modelos de regresión Logit bajo los enfoques clásico y bayesiano, con el objetivo de valorar la significancia estadística de las variables independientes en relación con la variable dependiente.
- En la tercera etapa, también con el conjunto de entrenamiento, se aplicaron los modelos de regresión Probit bajo los enfoques clásico y bayesiano, con el propósito de evaluar la significancia estadística de las variables independientes en correspondencia con la variable dependiente.
- En cuarto lugar, con la muestra de validación, se compararon los modelos estadísticos de regresión Logit y Probit bajo los enfoques clásico y bayesiano. Para evaluar el rendimiento de los modelos, se utilizó la matriz de confusión y sus métricas asociadas, tales como sensibilidad, precisión y F1-score. Posteriormente, se determinó el valor del área bajo la curva ROC y el índice de Gini, y se graficó la curva ROC para visualizar la capacidad discriminativa de cada modelo.
- Por último, utilizando el modelo de regresión y el enfoque más adecuado, se calculó el valor numérico del credit score a partir de las probabilidades de incumplimiento estimadas. Con estos valores, se construyeron deciles de credit score que permitieron segmentar a los solicitantes en niveles de riesgo crediticio: excelente, muy bueno, bueno, regular y malo. Finalmente, se propusieron recomendaciones de acciones para cada uno de los segmentos identificados.

2.11 SOFTWARE ESTADÍSTICO

Para el desarrollo y estimación de los modelos de regresión se utilizó el software RStudio Desktop, versión 2025.05.1-513, con soporte del lenguaje R, versión 4.5.1. Asimismo, se emplearon las bibliotecas caret, dplyr, ggplot2, MCMCpack y ROCR, que facilitaron la construcción, simulación, evaluación y visualización de los modelos estadísticos implementados.

3. DISCUSIÓN

3.1 División de la base de datos de clientes en entrenamiento y validación

La base de datos total de clientes del banco se dividió en una muestra de entrenamiento, compuesta por 3,918 clientes (70%) y una muestra de validación, compuesta por 1,666 clientes (30%).

3.2 Entrenamiento de los modelos de Regresión Logit bajo los enfoques clásico y bayesiano

Con la muestra de entrenamiento se aplicaron los modelos de regresión logit bajo los enfoques clásico y bayesiano, con el propósito de evaluar la significancia estadística de las variables independientes en correspondencia con la variable dependiente.

La estimación de los coeficientes para el modelo de regresión logit bajo el enfoque clásico se logra apreciar en la tabla 2.

Tabla 2

Resultados del Modelo de Regresión Logit bajo el enfoque clásico

Variables	Coefficientes	Pr(> z)
(Intercept)	-1.545751174	< 2e-16
int_rate	0.153002754	< 2e-16
emp_length	0.011446766	0.0259
annual_inc	-4.87252E-06	2.54E-08

Nota. Pr(>|z|) con un grado de significancia: $p < 0.05$.

La ecuación matemática del modelo de regresión logit bajo el enfoque clásico quedó expresada de la siguiente manera:

$$\text{Prob}(Y=1 | X)$$

$$=f(-1.545751+0.153003\text{int_rate}+0.011447\text{emp_length}-0.000005\text{annual_inc})=f(z)$$

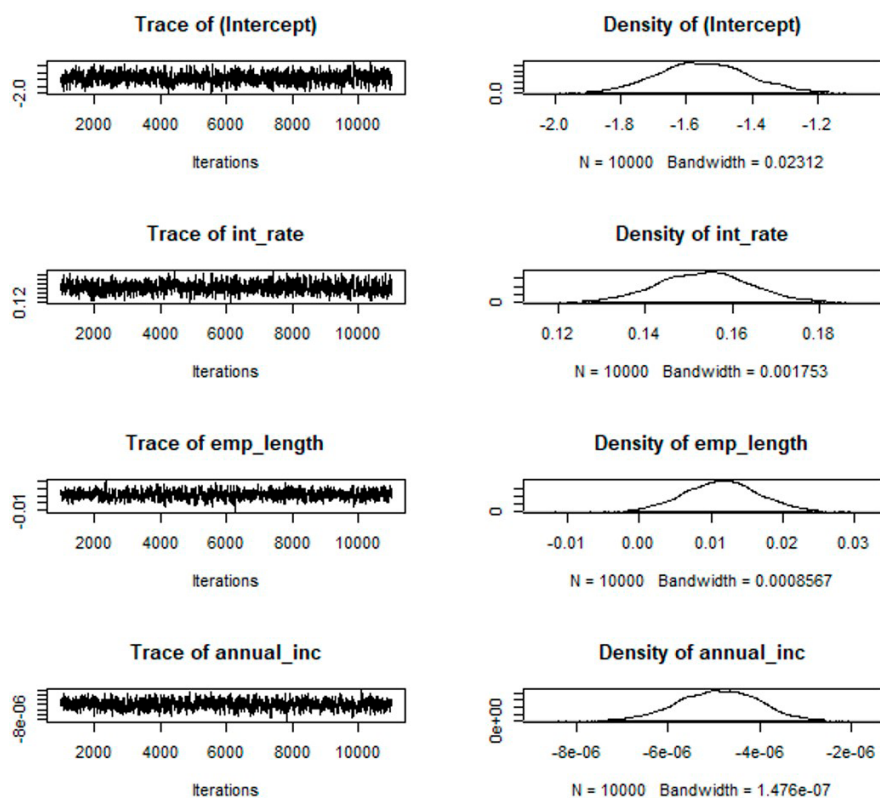
$$\text{Donde } f(z) = \frac{e^z}{1+e^z}$$

En la figura 2 se exhibe la estimación de los coeficientes del modelo estadístico de regresión logit bajo el enfoque bayesiano, obtenida mediante la simulación MCMC utilizando el algoritmo de Metropolis-Hastings. La figura muestra tanto las cadenas generadas por la simulación como las funcio-

nes de densidad posterior para cada uno de los coeficientes del modelo estadístico de regresión logit.

Figura 2

Estimación de los coeficientes del modelo de Regresión Logit bayesiano mediante MCMC



En la tabla 3 se presenta la media posterior de cada uno de coeficientes estimados del modelo estadístico de regresión logit desde la perspectiva bayesiana obtenida mediante la simulación MCMC utilizando el algoritmo de Metropolis-Hastings.

Tabla 3

Resultados del modelo de Regresión Logit desde la perspectiva bayesiano

Variables	Coefficientes
(Intercept)	-1.549872444
int_rate	0.153444736
emp_length	0.011564076
annual_inc	-4.9114E-06

La ecuación matemática del modelo de regresión logit desde la perspectiva bayesiano quedó expresada de la siguiente manera:

$$\text{Prob}(Y=1 | X)$$

$$= f(-1.549872 + 0.153445 \text{int_rate} + 0.011564 \text{emp_length} - 0.000005 \text{annual_inc}) = f(z)$$

$$\text{Donde } f(z) = \frac{e^z}{1+e^z}$$

3.3 Entrenamiento de los modelos de Regresión Probit bajo los enfoques clásico y bayesiano

Con la muestra de entrenamiento se aplicaron los modelos de regresión probit bajo los enfoques clásico y bayesiano, con el objetivo de evaluar la significancia estadística de las variables independientes en relación con la variable dependiente. La estimación de los coeficientes para el modelo de regresión probit bajo el enfoque clásico se presenta en la Tabla 4.

Tabla 4

Derivaciones del Modelo de Regresión Probit bajo el enfoque clásico

Variables	Coefficientes	Pr(> z)
(Intercept)	-0.973675778	< 2e-16
int_rate	0.094668264	< 2e-16
emp_length	0.006901243	0.0294
annual_inc	-2.72551E-06	1.80E-07

Nota. Pr(>|z|) con un grado de significancia: $p < 0.05$.

La ecuación matemática del modelo de regresión probit bajo el enfoque clásico quedó expresada de la siguiente manera:

$\text{Prob}(Y=1 | X)$

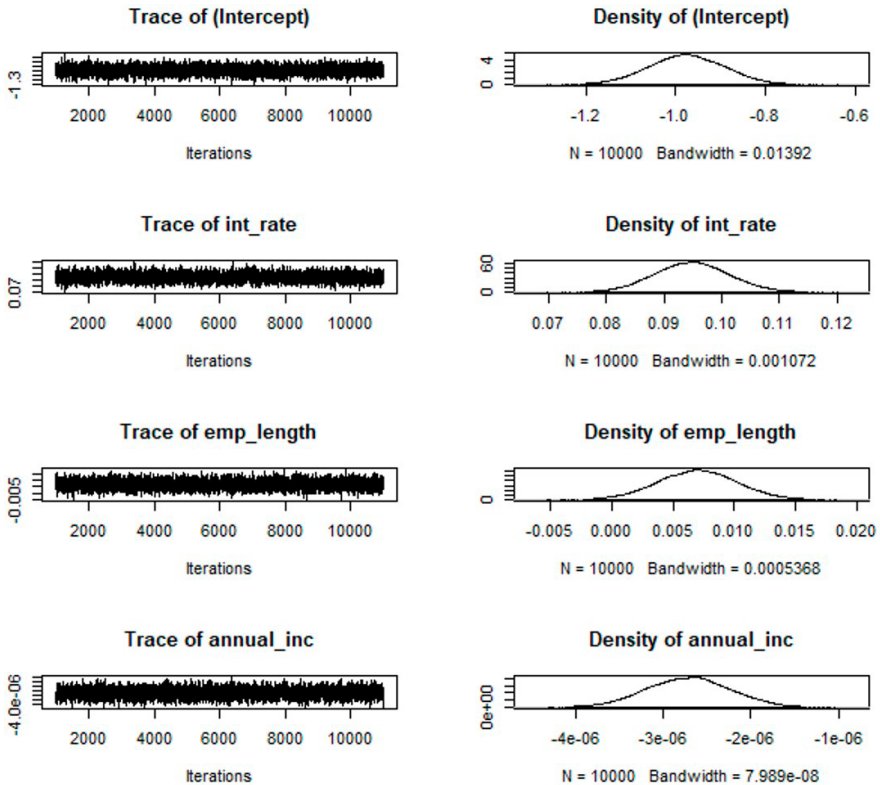
$$= f(-0.973676 + 0.094668 \text{int_rate} + 0.006901 \text{emp_length} - 0.000003 \text{annual_inc}) = f(z)$$

$$\text{Donde } f(z) = \int_{-\infty}^z \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{u^2}{2}} du$$

En la Figura 2 se presenta la estimación de los coeficientes del modelo estadístico de regresión probit bajo el enfoque bayesiano, obtenida mediante la simulación MCMC a través del muestreo de Gibbs. La figura muestra tanto las cadenas generadas por la simulación como las funciones de densidad posterior correspondientes a cada uno de los coeficientes del modelo.

Figura 3

Estimación de los coeficientes del Modelo de Regresión Probit bayesiano mediante MCMC



En la tabla 5 se presenta la media posterior de cada uno de coeficientes estimados del modelo estadístico de regresión probit bajo la perspectiva bayesiano obtenido mediante la simulación MCMC utilizando el muestreo de Gibbs.

Tabla 5
Resultados del Modelo de Regresión Probit bajo la perspectiva bayesiano

Variables	Coefficientes
(Intercept)	-0.976242658
int_rate	0.094804703
emp_length	0.006973918
annual_inc	-2.71173E-06

La ecuación matemática del modelo de regresión probit bajo la perspectiva bayesiano quedó expresada de la siguiente manera:

$$\text{Prob}(Y=1 | X)$$

$$= f(-0.976243 + 0.094805 \text{int_rate} + 0.006974 \text{emp_length} - 0.000003 \text{annual_inc}) = f(z)$$

$$\text{Donde } f(z) = \int_{-00}^z \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} e^{\left(-\frac{u^2}{2}\right)} du$$

3.4 Comparación de los modelos estadísticos de Regresión Logit y Probit bajo los enfoques clásico y bayesiano

En la Tabla 6 se presenta el desempeño comparativo de los modelos estadísticos de regresión logit y probit, evaluados tanto bajo el enfoque clásico como bajo el enfoque bayesiano. Para la comparación, se emplearon los indicadores estadísticos de sensibilidad, precisión, F1-score, área bajo la curva ROC y GINI, calculados a partir de la muestra de validación. Los resultados muestran que el modelo de regresión logit bajo el enfoque bayesiano alcanzó un desempeño predictivo superior en comparación con los demás modelos considerados.

Tabla 6

Comparación del desempeño de Modelos Logit y Probit, bajo los enfoques clásico y bayesiano

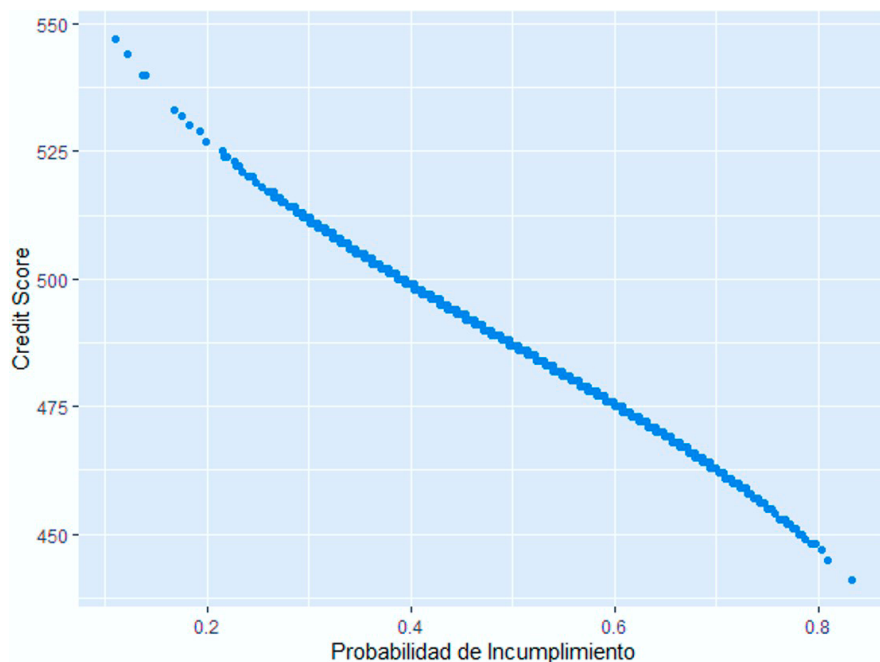
Modelos	Sensibilidad	Precisión	F1-score	ROC	GINI
Logit clásico	0.6423	0.6194	0.6307	65.6245	31.2489
Logit bayesiano	0.6423	0.6201	0.6310	65.6286	31.2573
Probit clásico	0.6388	0.6195	0.6290	65.4789	30.9578
Probit bayesiano	0.6388	0.6195	0.6290	65.4691	30.9382

3.5 Cálculo del Credit Score

A partir del modelo de regresión logit bajo la perspectiva bayesiana, se estimaron las probabilidades de incumplimiento y, con ellas, se calculó el valor del credit score. La Figura 4 muestra la correspondencia inversa entre la probabilidad de incumplimiento y el credit score; es decir, a mayor valor del credit score, menor será la probabilidad de incumplimiento, y viceversa.

Figura 4

Relación entre la Probabilidad de Incumplimiento y el Credit Score



Con los valores obtenidos del credit score, se procedió a segmentar a los solicitantes de préstamos personales en niveles de riesgo crediticio. En la Tabla 7 se presenta la distribución por deciles del credit score, indicando para cada decil el porcentaje de incumplimiento (default) y el valor del indicador de lift, calculado como la relación entre el porcentaje de incumplimiento de cada decil y el porcentaje de incumplimiento total de la muestra; esta información permitió identificar los deciles asociados a distintos niveles de riesgo, facilitando la categorización de los solicitantes y la implementación de estrategias diferenciadas de gestión del riesgo crediticio, tales como la asignación de tasas de interés, montos de préstamo y requerimientos de documentación en función del perfil de riesgo.

Tabla 7
Distribución por deciles del Credit Score

Decil	Loan status		% Default	LIFT
	0	1		
1	52	115	69%	1.37
2	70	97	58%	1.16
3	73	94	56%	1.12
4	59	108	65%	1.29
5	66	101	60%	1.21
6	81	86	51%	1.03
7	95	71	43%	0.85
8	98	68	41%	0.82
9	114	52	31%	0.62
10	122	44	27%	0.53
Total	830	836	50%	1.00

En la Tabla 8 se muestra la agrupación de los deciles del credit score en segmentos de riesgo crediticio, utilizando el indicador de lift como criterio de clasificación, donde valores superiores a 1 indican mayor riesgo de incumplimiento y valores inferiores a 1 menor riesgo; de esta manera, el segmento Malo corresponde al decil 1 (lift 1.37), el segmento Regular a los deciles 2, 3, 4 y 5 (lift entre 1.12 y 1.29), el segmento Bueno al decil 6 (lift 1.03), el segmento Muy Bueno a los deciles 7 y 8 (lift entre 0.82 y 0.85) y el segmento Excelente a los deciles 9 y 10 (lift entre 0.53 y 0.62); esta segmentación permite identificar claramente el nivel de riesgo de los solicitantes de préstamos personales y facilita la definición de políticas de

crédito y estrategias de gestión del riesgo adaptadas a cada categoría, tales como tasas de interés preferenciales y montos altos para los clientes excelentes, tasas favorables y evaluación de seguros para los muy buenos, tasas medias y verificación de ingresos para los buenos, tasas altas, documentación adicional y seguimiento para los regulares, y la no aprobación de préstamos para los clientes de menor riesgo crediticio.

Tabla 8

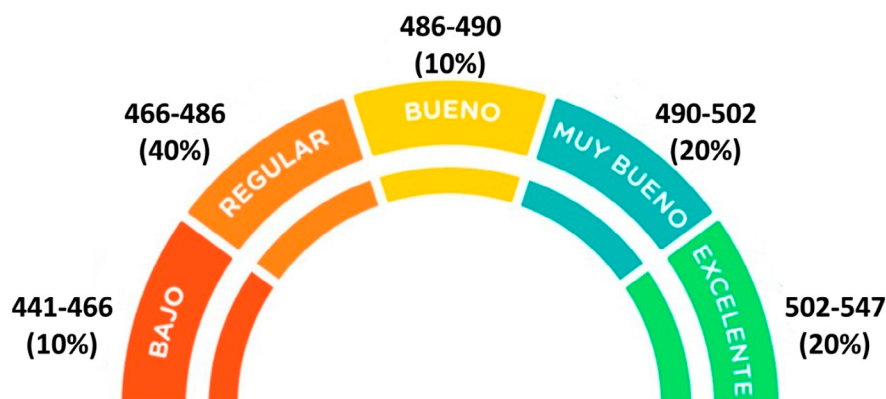
Segmentos de Riesgo a partir del Credit Score

Segmento	Loan status		% Default	LIFT
	0	1		
1.Malo	52	115	69%	1.37
2.Regular	268	400	60%	1.19
3.Bueno	81	86	51%	1.03
4.Muy Bueno	193	139	42%	0.83
5.Excelente	236	96	29%	0.58
Total	830	836	50%	1.00

Finalmente, se propusieron recomendaciones de acciones para cada uno de los segmentos de riesgo identificados. La Figura 5 muestra la distribución de los segmentos de credit score, incluyendo los valores mínimos y máximos correspondientes a cada categoría. Por ejemplo, el segmento de riesgo Excelente agrupa al 20% de los solicitantes de préstamos personales cuyos credit scores se encuentran en un rango de 502 a 547.

Figura 5

Segmentos de riesgo según el Credit Score



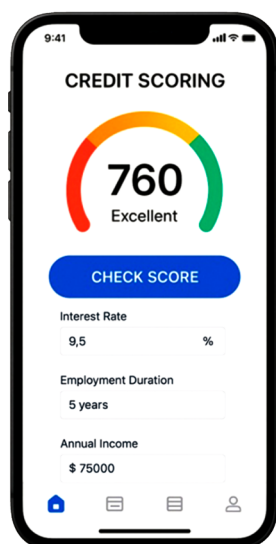
Las recomendaciones de acciones para cada uno de los segmentos de credit score de los solicitantes de préstamos personales son las siguientes:

- Segmento Excelente (credit score: 502 – 547): Se recomienda ofrecer préstamos con tasa de interés preferencial, montos altos y asignar un funcionario de negocio para brindar atención personalizada.
- Segmento Muy Bueno (credit score: 490 – 502): Se sugiere ofrecer préstamos con tasa de interés favorable, montos medios y evaluar la venta de un seguro adicional.
- Segmento Bueno (credit score: 486 – 490): Se recomienda otorgar préstamos con tasa de interés media, montos medios y solicitar documentación de ingresos de los últimos 6 meses.
- Segmento Regular (credit score: 466 – 486): Se sugiere ofrecer préstamos con tasa de interés alta, montos bajos, solicitar documentación de ingresos de los últimos 12 meses, requerir un aval y realizar un seguimiento del cumplimiento del cliente durante los meses siguientes.
- Segmento Bajo (credit score: 441 – 466): Se recomienda rechazar la solicitud de préstamo personal debido al alto riesgo de incumplimiento.

Finalmente, se propone el diseño de una aplicación móvil que integre y ponga en producción el modelo estadístico de regresión logística bajo el enfoque bayesiano. Para ello, se sugiere la implementación de una API en la nube que permita el ingreso de las variables `int_rate`, `emp_length` y `annual_inc`, garantizando así la evaluación en tiempo real de los solicitantes de préstamos personales. La Figura 6 presenta el diseño conceptual de la aplicación móvil para la valoración del credit scoring.

Figura 6

Diseño de Aplicación Móvil para la evaluación de Credit Scoring



4. DISCUSIÓN

Las derivaciones obtenidas en esta investigación confirman la utilidad del credit scoring como un instrumento estadístico clave para la gestión del riesgo crediticio. La segmentación de los solicitantes de préstamos personales en función del credit score, calculado a partir de sus probabilidades de incumplimiento, permite a las entidades financieras optimizar sus políticas de otorgamiento de crédito, adaptándolas al perfil de riesgo de cada peticionario. Este hallazgo coincide con estudios previos, como los de Dassatti (2019), Nopper. (2020) y Reyes y Sosa (2022), quienes destacan que un sistema de puntuación crediticia adecuado contribuye a reducir la probabilidad de pérdidas financieras en una entidad.

En particular, el análisis comparativo entre los modelos estadísticos de regresión logit y probit, estimados bajo los enfoques clásico y bayesiano, permitió identificar diferencias significativas en su capacidad predictiva. Según los hallazgos obtenidos, el modelo logit bayesiano presentó el mejor desempeño, con indicadores de precisión de 0.6201, F1-score de 0.6310, área bajo la curva ROC de 65.63 e índice de GINI de 31.26. Estos resultados sugieren que la incorporación de información a priori, característica del enfoque bayesiano, mejora

notablemente la estimación de los coeficientes y, por ende, la capacidad predictiva del modelo.

Asimismo, el uso del enfoque bayesiano, particularmente mediante la simulación Monte Carlo vía Cadenas de Markov (MCMC) y el algoritmo de Metropolis-Hastings, permitió superar el desempeño del enfoque clásico. La evidencia empírica respalda también la elección del modelo logit frente al probit, en línea con la literatura que indica que, aunque ambos modelos suelen generar resultados similares, el logit ofrece mayor estabilidad y facilidad de interpretación en aplicaciones prácticas de riesgo crediticio (Hosmer et al., 2013).

En síntesis, los hallazgos aportan evidencia sólida sobre la superioridad del modelo logit bajo perspectiva bayesiana para el desarrollo de un modelo de credit scoring, alineándose con lo reportado por Webster (2011), quien enfatiza la relevancia de incorporar distribuciones a priori para optimizar la tipificación de clientes con alta probabilidad de incumplimiento. Estos resultados también son consistentes con las evaluaciones realizadas por Agrawal et al. (2021) y Mousavi y Gazori (2023).

Las implicancias son positivas para la toma de decisiones en instituciones financieras, que pueden mejorar la gestión del riesgo crediticio mediante modelos estadísticos basados en enfoques bayesianos, tal como sugieren Banca & Economía (2022) y Hilbck (2022).

Entre las limitaciones del estudio se encuentra el diseño transversal, que impide observar cambios en el comportamiento de los solicitantes de crédito a lo largo del tiempo, dado que el riesgo crediticio es dinámico y puede verse influido por factores macroeconómicos no capturados en la investigación. Otra limitación relevante es que la base de datos proviene de la plataforma DataCamp, diseñada para fines académicos; aunque los datos reflejan características reales de solicitantes de crédito, los resultados podrían mejorar su precisión si se aplicara el modelo con información de un banco en actividad.

5. CONCLUSIONES

Se evaluó el desempeño del modelo estadístico de regresión logit bajo los enfoques clásico y bayesiano para predecir la probabilidad de incumplimiento de solicitantes de préstamos personales. En el enfoque clásico, se obtuvieron los siguientes resultados: sensibilidad de 0.6423, precisión de 0.6194, F1-score de 0.6307, área bajo la curva ROC de 65.6245 e índice de Gini de 31.2489. En el

enfoque bayesiano, se mantuvo la misma sensibilidad de 0.6423, presentándose ligeras mejoras en la precisión (0.6291), F1-score (0.6310), área bajo la curva ROC (65.6286) e índice de Gini (31.2573). Estos resultados indican que ambos enfoques para el modelo logit muestran un desempeño predictivo muy similar en la discriminación del riesgo crediticio, aunque el enfoque bayesiano ofrece una leve ventaja.

Asimismo, se evaluó el modelo estadístico de regresión probit bajo los mismos enfoques. En el enfoque clásico, se obtuvieron: sensibilidad de 0.6388, precisión de 0.6195, F1-score de 0.6290, área bajo la curva ROC de 65.4789 e índice de Gini de 30.9578. En el enfoque bayesiano, los resultados fueron idénticos para la sensibilidad, precisión y F1-score, mientras que se observaron ligeras disminuciones en el área bajo la curva ROC (65.4691) y en el índice de Gini (30.9382). Esto evidencia que, para el modelo probit, ambos enfoques presentan un desempeño prácticamente equivalente en la discriminación del riesgo crediticio.

Al comparar los modelos logit y probit bajo los enfoques clásico y bayesiano para el desarrollo de un modelo de credit scoring, se observó que, según el indicador de sensibilidad, el modelo logit (clásico y bayesiano) obtuvo un valor de 0.6423, superior al del probit (0.6388). Sin embargo, considerando los indicadores de precisión, F1-score, área bajo la curva ROC e índice de Gini, el modelo logit bajo el enfoque bayesiano presentó el mejor desempeño, con valores de 0.6201, 0.6310, 65.6286 y 31.2573, respectivamente. Por lo tanto, se concluye que este modelo es el más adecuado para segmentar a los solicitantes y proponer acciones según su nivel de riesgo mediante un sistema de credit scoring.

Finalmente, aunque las diferencias numéricas entre los enfoques clásico y bayesiano, tanto para logit como probit, son pequeñas, los resultados sugieren que el modelo Logit Bayesiano ofrece un desempeño más consistente y ligeramente superior. Esto puede ser relevante en contextos de credit scoring, donde incluso incrementos marginales en la capacidad predictiva pueden traducirse en decisiones más eficientes de segmentación y gestión del riesgo crediticio.

6. RECOMENDACIONES

Como parte del impacto de la investigación en las políticas públicas, se recomienda que la SBS considere los resultados de este estudio como una he-

herramienta estadística basada en modelos bayesianos para promover prácticas más eficientes y responsables en el sistema crediticio. La aplicación de estos modelos puede contribuir a la estabilidad del sistema financiero, al facilitar una mejor segmentación del riesgo y respaldar la implementación de tasas de interés reguladas, así como programas de acompañamiento financiero orientados a prevenir el sobreendeudamiento de los solicitantes. De esta manera, los hallazgos fortalecen la articulación entre la modelación estadística y el diseño de políticas públicas, favoreciendo una toma de decisiones basada en evidencia.

Asimismo, se recomienda a los tomadores de decisiones del sector financiero utilizar los resultados de esta investigación como insumo técnico para mejorar la calidad de la cartera crediticia. Esto incluye aplicar criterios objetivos para la aprobación de préstamos personales, considerando variables clave como el monto solicitado, el plazo y la tasa de interés, en función del perfil de riesgo crediticio de cada solicitante. De este modo, se contribuye a una gestión más eficiente y responsable del riesgo de crédito.

REFERENCIAS

- Adams, N. M., & Hand, D. J. (1999). Comparing classifiers when the misallocation costs are uncertain. *Pattern Recognition*, 32(7), 1139–1147. Recuperado de [https://doi.org/10.1016/S0031-3203\(98\)00154-X](https://doi.org/10.1016/S0031-3203(98)00154-X)
- Agrawal, S., Ahirao, P., Kumar, S., & Dere, P. (2021). Credit Score Evaluation of Customer Using Machine Learning Algorithms. *Proceedings of the 4th International Conference on Advances in Science & Technology (ICAST2021)*. <http://dx.doi.org/10.2139/ssrn.3867420>
- Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422), 669–679. <https://doi.org/10.2307/2290350>
- Banca & Economía. (2022). *Uso de información alternativa para fortalecer los modelos de scoring*. Asobancaria. <https://www.asobancaria.com/ws/semanas-economicas/1346-BE.pdf>
- Barrios, J. I. (2019). *La matriz de confusión y sus métricas*. Heath Big Data. <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>
- Dassatti, C. (2019). *Modelos de Score Crediticio: revisión metodológica y análisis a partir de datos de encuesta* (Credit Score Models: Methodological Review and Analysis Based on Survey Data). Available at SSRN 3443515. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3443515

- Espín, O., & Rodríguez, C. (2013). Metodología para un scoring de clientes sin referencias crediticias. *Cuadernos de Economía*, 32(59), 139-165. Recuperado de <https://repositorio.unal.edu.co/handle/unal/72537>
- Fonseca, P. G., & Lopes, H. D. (2017). *Calibration of machine learning classifiers for probability of default modelling*. arXiv. <https://doi.org/10.48550/arXiv.1710.08901>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press. <https://doi.org/10.1201/9780429258411>
- Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University. <https://n9.cl/pbfe15>
- Hand, D. J., & Anagnostopoulos, C. (2013). When is the area under the receiver operating characteristic curve an appropriate measure of classifier performance? *Pattern Recognition Letters*, 34(5), 492-495. <https://doi.org/10.1016/j.patrec.2012.12.004>
- Hilbck, M. (2022). *Optimización para la toma de decisiones: un enfoque analítico*. UCP. <https://repositorio.pucp.edu.pe/index/handle/123456789/187861>
- Macskassy, S., & Provost, F. (2003). Confidence Bands for Roc Curves. NYU Stern School of Business. <https://core.ac.uk/reader/43020236>
- Martin, A. D., Quinn, K. M., & Park, J. H. (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software*, 42(9), 1-21. <https://doi.org/10.18637/jss.v042.i09>
- Moretz, B. DataCamp. *Repositorio GitHub*. GitHub <https://github.com/bmoretz/DataCamp/tree/master/credit.risk/data>
- Mousavi, S. F., & Gazori, N. A. (2023). A review of credit rating models: A combined analysis and suggestions for future research. *Journal of Industrial and Systems Engineering*, 15(1), 306-327. <https://n9.cl/u43n0>
- Nopper, T. K. (2020). *Alternative data and the future of credit scoring*. Data for Progress. <https://www.filesforprogress.org/memos/alternative-data-future-credit-scoring.pdf>
- Reyes, M. M. A., & Sosa, M. (2022). Modelo de puntuación crediticia para tarjeta de crédito en México: una aproximación logística. *Ensayos. Revista de economía*, 41(1), 17-52. <https://doi.org/10.29105/ensayos41.1-2>
- Thomas, L. C., Crook, J. N., & Edelman, D. B. (2002). Credit scoring and its applications. SIAM. <https://epubs.siam.org/doi/pdf/10.1137/1.9781611974560.bm>
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. John Wiley & Sons. <https://n9.cl/uqmvq>
- Wooldridge, J. (2010) *Introducción a la Econometría*. (4ª ed.) Cengage Learning. <https://n9.cl/2ukss>