

Evaluación de modelos Red Neuronal Perceptrón Multicapa y Logit para la clasificación de impago en préstamos bancarios.

Walter Jeremías López Flores*

Resumen.

El objetivo del estudio es evaluar el desempeño de redes neuronales como técnicas modernas para clasificar el riesgo de *default*, contra métodos estadísticos tradicionales Logit. Los datos provienen de la cartera de préstamos personales de un banco de Honduras. Los resultados con datos desequilibrados demuestran un desempeño superior de las redes. Al aplicar sobremuestreo, no hubo mayor impacto estas, pero sí en las regresiones, donde sus nuevos resultados casi las alcanzan en el modelo final seleccionado, el cual presenta una exactitud de 99.16%, puntuación F1 del 98.26% y curvas ROC y PR con AUC de 98.68% y 97.69% respectivamente.

Palabras clave: Aprendizaje de Máquina Supervisado, Red Neuronal Perceptrón Multicapa, Regresión Logística, Riesgo de crédito, Impago de préstamos.

Clasificación JEL: C01, C45, C51, C52, G21.

Evaluation of Multilayer Perceptron Neural Network and Logit models for the classification of default in bank loans.

Abstract.

The objective of the study is to evaluate the performance of neural networks as modern techniques to classify the risk of default, against traditional Logit statistical methods. The data comes from the personal loan portfolio of a Honduran bank. Results with unbalanced data demonstrate superior network performance. When applying oversampling, there was no greater impact on these, but there was on the regressions, where their new results almost reached them in the final selected model, which presents an accuracy of 99.16%, F1-score of 98.26% and ROC and PR curves with AUC of 98.68% and 97.69% respectively.

Keywords: Supervised Machine Learning, Multilayer Perceptron Neural Network, Logistic Regression, Loans default, Credit risk.

JEL classification: C01, C45, C51, C52, G21.

* Profesor de la Facultad de Postgrado en Universidad Tecnológica Centroamericana, Honduras. Ph.D. candidate in Economía y Empresa. Email: wjlopez@unitec.edu. ORCID: 0000-0002-7874-796X.

1. INTRODUCCIÓN.

Los bancos cuentan con una gran cantidad de datos sobre sus prestatarios, la cual se puede emplear para predecir la probabilidad de que estos incumplan o no con sus préstamos. Lo anterior, según Tiwari (2018) se debe a que la mayor parte de la población mundial tiene acceso a servicios bancarios, la multiplicación de los consumidores en los últimos años y al mismo tiempo el avance tecnológico ha permitido recopilar y guardar grandes datos que representan el comportamiento del cliente y los riesgos en torno al préstamo. Dada la importancia del riesgo crediticio, varios estudios han propuesto adoptar en los bancos modelos estadísticos para mejorar su evaluación y aumentar la precisión en la predicción de los modelos existentes (Obare, Njoroge y Muraya, 2019).

Las decisiones de evaluación sobre riesgo crediticio son determinantes clave para el éxito de las instituciones financieras, debido a las grandes pérdidas que resultan en la industria crediticia por decisiones equivocadas (Lahsasna, Ainon y Wah, 2010). Es por ello que la evaluación del riesgo de crédito es esencial antes de tomar cualquier decisión de préstamo, por lo que el desempeño adecuado del departamento de crédito garantiza la rentabilidad y estabilidad de un banco al examinar el historial y antecedentes financieros de los clientes para tomar decisiones crediticias, siendo este un factor determinante clave para reducir el riesgo de crédito (Bekhet y Eletter, 2014).

La evaluación de este tipo de riesgo se basa en la probabilidad de que el prestatario incumpla con sus obligaciones, situación que en el argot financiero se llama *default* (Saavedra y Saavedra, 2010) o impago. Al producirse impagos en los préstamos se genera tensión económica, ya que a los prestatarios de calidad se les niega el acceso al crédito que pueden utilizar para desarrollar la economía, y dichos incumplimientos hacen que las instituciones financieras incurran en pérdidas tanto de capital como de intereses (Obare y Muraya, 2018) y cuando se rechaza el préstamo, existe el riesgo de perder un cliente potencialmente rentable frente a la competencia y el riesgo del costo de oportunidad (Bekhet y Eletter, 2014).

La minería de datos es un área prometedora del análisis de datos cuyo objetivo es extraer conocimiento útil de una enorme cantidad de conjuntos de datos complejos, usando métodos estadísticos, especialmente técnicas de aprendizaje automático o de máquina (ML por sus siglas en inglés) para modelar y predecir pérdidas el incumplimiento de préstamos, siendo algunos algoritmos adecuados para esta tarea: la regresión logística, k -vecinos más cercanos (k -NN), árboles de clasificación y regresión (CART) y bosques aleatorios (RF), siendo la regresión logística la técnica más ampliamente utilizada para fines de clasificación (Tiwari, 2018) desde la publicación seminal de Ohlson (1980) donde aplicó el modelo Logit para la previsión de insolvencias, abriendo la posibilidad de utilizar otras metodologías en este campo de investigación, siendo una década después cuando

Odom y Sharda (1990) incorporaron la aplicación de técnicas basadas en Inteligencia Artificial (IA) conocidas como Redes Neuronales Artificiales (ANN por sus siglas en inglés) (do Prado et al., 2019).

Lahsasna et al. (2010) explican que las ANN son representaciones matemáticas inspiradas en el funcionamiento del cerebro humano, compuestas por una serie de neuronas que trabajan en paralelo sin ningún control centralizado, y dispuestas en una estructura de capas mediante un sistema de conexiones ponderadas que determinan el flujo de información a través de la red para modelar problemas complejos del mundo real, las cuales se han establecido como una alternativa seria a los modelos estadísticos tradicionales en varias disciplinas, ya que muchos estudios han concluido que estas superaron a los métodos estadísticos en términos de precisión de clasificación; mientras que Tsai y Wu (2008) aseguran que la red de tipo Perceptrón Multicapa (MLP por sus siglas en inglés) entrenada por el algoritmo de *backpropagation* (retropropagación) es la técnica más utilizada para los problemas de toma de decisiones financieras y West (2000) que es la arquitectura más frecuente en aplicaciones comerciales, incluyendo la clasificación crediticia.

En base a lo anterior se ha propuesto como objetivo principal de este estudio, evaluar el desempeño entre estos nuevos modelos para clasificar a los clientes que tienen préstamos vigentes y/o que solicitan nuevas aplicaciones de crédito, si pueden o no caer en el impago de los mismos, tomando como caso de estudio la cartera de préstamos de un banco hondureño; para lo cual, se realizarán varios experimentos con clasificadores de tipo MLP comparándolos contra equivalentes Logit, por ser modelos estadísticos tradicionales ampliamente utilizado en la literatura para modelar el impago de créditos, y de esa manera analizar el desempeño entre estos para realizar dicha clasificación binaria.

Por otro lado, los datos del mundo real a menudo son ruidosos e incompletos (Golbayani, Wang y Florescu, 2020) y un problema común dentro del ML son los conjuntos de datos desequilibrados, puesto que la mayoría de los modelos de clasificación tradicionales no pueden manejarlos, dado el alto costo de clasificación erróneo en la clase minoritaria; problemática que se aborda con una técnica llamada sobremuestreo, la cual permite crear una muestra de clase específica para que la distribución de clases del conjunto original pueda equilibrarse (Jiang, Zhang y Zou, 2019).

En vista de ello, también se propone en este estudio analizar la manera en que mejora el desempeño de ambos clasificadores al entrenar con datos balanceados en sus clases de salida al aplicar sobremuestreo.

2. MARCO TEÓRICO.

El riesgo de crédito es la probabilidad de que una entidad no haga frente, parcial o totalmente, a su obligación de devolver una deuda o rendimiento convenido a su vencimiento sobre un instrumento financiero, ya sea por quiebra, iliquidez u otra razón (Saavedra y Saavedra, 2010).

Dicho problema es abordado según Lahsasna et al. (2008) mediante modelos de calificación crediticia, basados en métodos estadísticos o de investigación operativa, los cuales se crean a partir de información histórica de miles de clientes reales; para cada solicitud se toma un formulario y un historial durante un período fijo y luego se decide sobre si el cliente es bueno o malo, por lo que se encuentra en el dominio de problemas de clasificación, capturando la relación entre la información histórica y el desempeño crediticio a futuro a través de la siguiente función:

$$f(x_1, x_2, \dots, x_m) = y_n \quad (1)$$

Donde cada cliente contiene m atributos o características: x_1, x_2, \dots, x_m , mientras que cada y_i denota el tipo de cliente (si cae en impago o no, que es el caso de clasificación binaria de este estudio) y f es la función o modelo que mapea entre las características del cliente (entradas) y su solvencia (salida), la tarea del modelo es predecir el valor de y , predicción que se abordará con los modelos ANN de tipo MLP y Logit.

En cuanto a las variables independientes o características consideradas en estudios previos para explicar el impago, Salazar (2013) resalta las siguientes como los principales determinantes del comportamiento crediticio: estado civil, nacionalidad, sexo, número de hijos, edad, tenencia de teléfono, tiempo en el domicilio actual, región geográfica, profesión, sector de actividad económica, tiempo en el trabajo actual, ingresos y gastos mensuales, propiedad de la vivienda, créditos concedidos con anterioridad, cantidad, duración y destino del préstamo.

2.1. Las Redes Neuronales Perceptrón Multicapa en la clasificación crediticia.

Khashman (2010) considera que el uso de las ANN para la calificación y evaluación de créditos ha sido efectivo durante la última década, dada su capacidad en tales aplicaciones, la cual se debe a la forma en que opera la red y la disponibilidad de datos de entrenamiento, lo cual es más evidente cuando se utilizan redes MLP basadas en el algoritmo de aprendizaje de retropropagación, ilustradas en la Figura 1. Esta es la arquitectura ANN más clásica, conocida también como *Feed Forward* (alimentación hacia adelante) y la más frecuente para problemas de calificación crediticia, ya que permite explorar tendencias complejas no lineales en los datos (Golbayani et al., 2020).

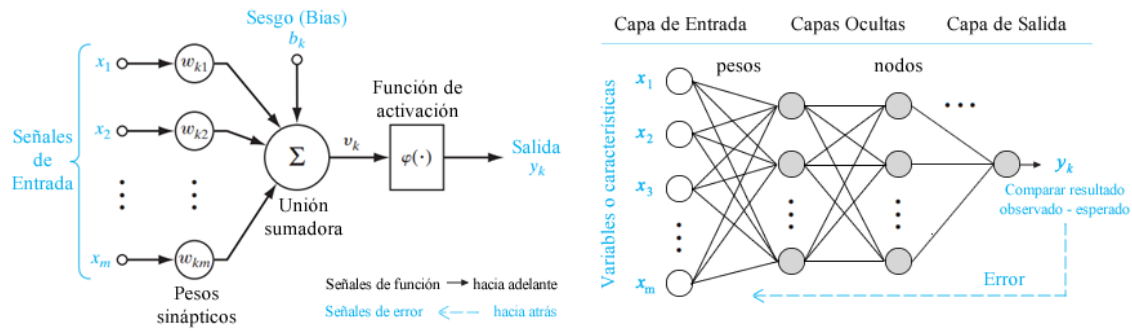


Figura 1. Arquitectura de una Red *Feed Forward* o Perceptrón Multicapa.

Fuente: adaptado y traducido por el autor a partir de Haykin (2009).

Cada nodo de la red es una unidad de procesamiento de información fundamental llamado neurona o perceptrón, consiste en un clasificador lineal que separa dos clases y en términos matemáticos una red se describe de la siguiente manera (Haykin, 2009; Golbayani, Florescu y Chatterjee, 2020):

$$y_k = \varphi \left(\sum_{i=1}^m w_{ki} x_i + b_k \right) = \varphi(w^T x + b) \quad (2)$$

Donde x_1, x_2, \dots, x_m son las señales de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ son los respectivos pesos sinápticos de la neurona k , los cuales pueden ser vistos como vectores; b_k es el sesgo (*bias*) cuyo efecto es aplicar una transformación afinada a la salida; $\varphi(\cdot)$ es la función de activación y y_k es la salida de la neurona. En la Figura 2 se muestran las funciones de activación más comúnmente utilizadas en redes neuronales.

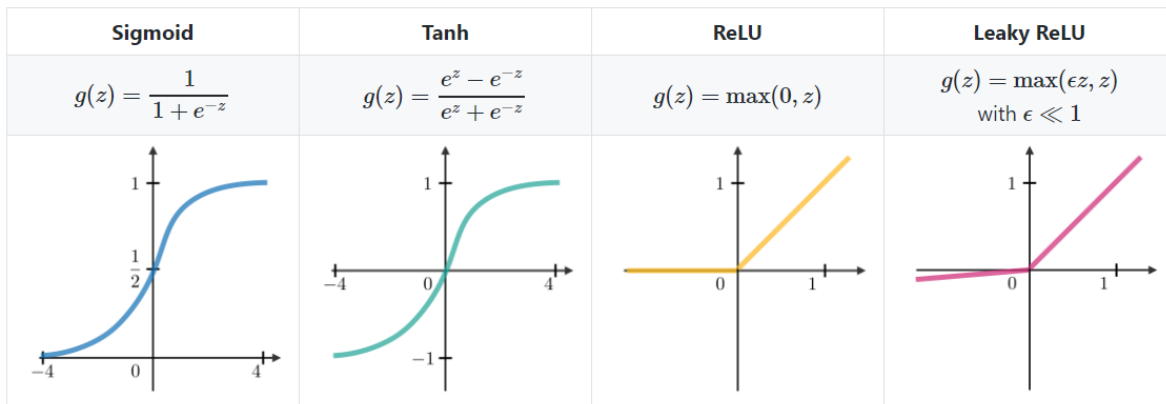


Figura 2. Funciones de activación más comúnmente utilizadas en redes neuronales.

Fuente: Amidi y Amidi (s.f.).

Brownlee (2021) define las funciones de activación o de transferencia como la suma ponderada de la entrada que se transforma en una salida en una capa de la red, y propone la siguiente metodología para selección de funciones: para las capas ocultas, en el caso de las redes MLP o convolucionales usar ReLU, y para redes recurrentes usar *sigmoid* (sigmoidea) o Tanh (Hiperbólica Tangente); luego, para la capa de salida, para los problemas de clasificación binaria sugiere utilizar sigmoidea. Usar ReLU asegura que la salida nunca sea negativa, dado que si $z > 0$ la salida sigue siendo z , y si $z < 0$ la salida es 0 (Babs, 2018).

Como se ilustra en la Figura 1, las redes MLP están compuestas por varios perceptrones organizados en varias capas, divididas en tres grupos: una capa de entrada (*input layer*), una o varias ocultas (*hidden layers*) y una capa de salida (*output layer*), constan de: pesos entre sus conexiones (cuanto mayor sea el valor mayor importancia tendrá la neurona), la función de activación, el valor para el sesgo y el cálculo de error entre la salida deseada y la real para ajustar los parámetros, los pesos y el sesgo (que captura los factores imprevistos o no observables) de forma inversa (*backpropagation*) durante el entrenamiento para minimizar el error, el cual se representa de la siguiente manera (Peña y Orellana, 2018; Babs, 2018; Golbayani et al., 2020):

$$\Delta w = \eta * e_k * x_{jk} \quad (3)$$

Donde e_k representa el error en la unidad de salida k y x_{jk} es la entrada que causa el error, η es la tasa de aprendizaje que define que tanto cambia el peso para corregir el error, estos son propagados de manera inversa a la red y el cambio de los pesos en la capa oculta sigue la fórmula (3) a excepción del término de error, que se calcula con la derivada parcial del error:

$$e_j = w_j * e_k * d(y_k) \quad (4)$$

Donde d es la derivada de transferencia. La red se mantiene haciendo esto hacia adelante y atrás hasta que el error no se pueda disminuir más, dicho estado se llama convergencia (Golbayani et al., 2020). Amidi y Amidi (s.f.) explican que la derivada con respecto al peso w es calculada utilizando la regla de la cadena de la siguiente forma:

$$\frac{\partial L(z, y)}{\partial w} = \frac{\partial L(z, y)}{\partial a} \times \frac{\partial a}{\partial z} \times \frac{\partial z}{\partial w} \quad (5)$$

Como resultado, el peso se actualiza de la siguiente manera:

$$w \leftarrow w - \alpha \frac{\partial L(z, y)}{\partial w} \quad (6)$$

La tasa de aprendizaje, también denotada por α indica a qué ritmo se actualizan las ponderaciones, se puede cambiar de forma adaptativa, siendo *Adam* el método más popular en la actualidad (Amidi y Amidi, s.f.), el cual es un algoritmo de optimización basado en gradientes de primer orden de funciones objetivas estocásticas, y en estimaciones adaptativas de momentos de orden inferior, sencillo de implementar, computacionalmente eficiente, con pocos requisitos de memoria, invariante al cambio de escala diagonal de los gradientes y adecuado para problemas con grandes cantidades de datos o parámetros, así como objetivos no estacionarios y problemas con gradientes muy ruidosos o dispersos (Kingma y Ba, 2017).

Entre los tipos de aprendizaje automático o de máquina (ML por sus siglas en inglés) están: supervisado, no supervisado y por refuerzo (Shin y Lee, 2020), cuando las redes son supervisadas, es requerido que el usuario indique la salida deseada, a partir de ello, aprende a detectar la relación entre las entradas y salidas suministradas mediante el proceso adaptativo e iterativo explicado; una vez entrenada, se le presentan nuevos datos, que no ha visto antes y se prueba para comprobar la bondad del conjunto de pesos, y al ofrecer un rendimiento óptimo se considera lista para trabajar (Pérez y Fernández, 2007). Las funciones de pérdidas $L: (z, y) \in \mathbb{R} \times Y \mapsto L(z, y) \in \mathbb{R}$ toman como entradas el valor pronosticado z correspondiente al valor de datos real y para mostrar qué tan diferentes son (Amidi y Amidi, s.f.), en la Figura 3 se muestran las más comunes.

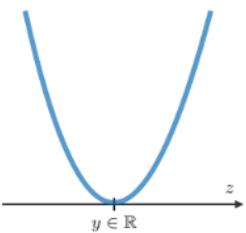
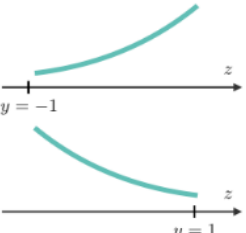
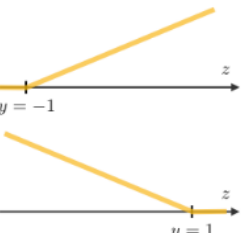
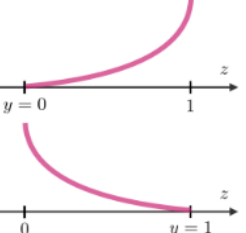
Least squared error	Logistic loss	Hinge loss	Cross-entropy
$\frac{1}{2}(y - z)^2$	$\log(1 + \exp(-yz))$	$\max(0, 1 - yz)$	$-\left[y \log(z) + (1 - y) \log(1 - z)\right]$
			
Linear regression	Logistic regression	SVM	Neural Network

Figura 3. Funciones de pérdida más comúnmente utilizadas en redes neuronales.

Fuente: Amidi y Amidi (s.f.).

Para calcular la pérdida entre etiquetas verdaderas y predichas cuando solo hay dos clases de etiquetas (0 y 1), se utiliza una función probabilística llamada *binary crossentropy* (entropía cruzada binaria) (Keras, s.f.). Mientras que la función de costo J se utiliza comúnmente para medir el desempeño de un modelo, y se define con la función de pérdida L de la siguiente manera (Amidi y Amidi, s.f.):

$$J(\theta) = \sum_{i=1}^m L(h_{\theta}(x^{(i)}), y^{(i)}) \quad (7)$$

Al señalar que la tasa de aprendizaje $\alpha \in \mathbb{R}$, la regla de actualización para el descenso del gradiente (*gradient descent*) se expresa con la tasa de aprendizaje y la función de costo J así:

$$\theta \leftarrow \theta - \alpha \nabla J(\theta) \quad (8)$$

Para evitar el ajuste excesivo u *overfitting* (sobre-entrenamiento) en la red, se utilizan métodos de regularización estocástica, como el *dropout* (abandono) dada su simplicidad (Kingma y Ba, 2017), esta capa de abandono, establece de forma aleatoria las unidades de entrada en 0 con una frecuencia de velocidad a cada paso durante el entrenamiento, lo cual ayuda a evitar sobreajuste, y las que no están configuradas en 0 se escalan en $1/(1 - \text{tasa})$ de manera que la suma de todas las entradas no cambia (Keras, s.f.).

Tsai y Wu (2008) realizan un análisis entre nueve trabajos relacionados con el uso de modelos MLP para la predicción de quiebras y calificación crediticia (Fan y Palaniswami, 2000; West, 2000; Atiya, 2001; Lee et al., 2002; Huang et al., 2004; Min y Lee, 2005; Shin et al., 2005; West et al., 2005; Lee et al., 2006) en el que se usan datos de créditos bancarios de Australia, Alemania, Estados Unidos, Taiwán y Corea del Sur; su principal conclusión es que la mayoría de investigaciones en este campo utilizan un conjunto de datos para validación y solo examinan el rendimiento promedio de predicción y la exactitud de sus modelos sin considerar los errores de Tipo I y Tipo II, y muy pocos consideran emplear múltiples clasificadores.

A pesar de que la tasa de clasificación de los modelos ANN y su exactitud es alta, son criticados por su falta de capacidad explicativa dada su naturaleza de caja negra, siendo esta su principal deficiencia en el ámbito empresarial (Lahsasna et al., 2010), además de otras fallas como: propensión a quedar atrapada en un óptimo local, tiempo de entrenamiento costoso, sobreajuste y requerimiento de grandes instancias de aprendizaje (Obare et al., 2019).

2.2. El modelo Logit para clasificación crediticia.

La regresión logística es una técnica prestada por el ML del campo de las estadísticas (Bronwlee, 2016). Salazar (2013) explica que este modelo pertenece al grupo de regresión de respuesta cualitativa, donde la variable dependiente toma diversos valores según una clasificación previamente establecida, asignando generalmente valores de 0 o 1 para indicar la presencia o ausencia de un atributo, en este caso: el incumplimiento en el pago del préstamo; y que la regresión de esta variable cualitativa, frente a un grupo de

variables exógenas permite hacer un análisis de probabilidad para el que se utiliza una función de distribución acumulativa logística, donde la probabilidad condicional está dada por:

$$P_i = \Pr(Y = 1|X_i) = \int_{-\infty}^{\beta'x} \varphi(z)dz = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}} = \frac{1}{1 + e^{-Z_i}} = \frac{e^{Z_i}}{1 + e^{Z_i}} \quad (9)$$

Mientras que la probabilidad de que no suceda el evento está dada por:

$$1 - P_i = \frac{1}{1 + e^{Z_i}} \quad (10)$$

Los modelos Logit usan la función sigmoidea, representada en la Figura 2. Brownlee (2016) explica que los valores de entrada x , se combinan linealmente utilizando pesos o valores de coeficientes (denominados Beta) para predecir un valor de salida y binario de la siguiente manera:

$$\hat{y} = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad (11)$$

Donde \hat{y} es la salida prevista, β_0 es el sesgo o término de intersección y β_1 es el coeficiente para el valor de entrada x . Cada columna de los datos de entrada tiene un coeficiente β asociado, que es un valor real constante que el modelo debe aprender de los datos de entrenamiento, usando la estimación de máxima verosimilitud (*maximum-likelihood*) cuyo algoritmo de aprendizaje busca valores para los coeficientes que minimizan el error en las probabilidades predichas por el modelo con respecto a los datos.

En estudio de Desai et al. (1996) compararon varios modelos usando regresión logística análisis discriminante y redes MLP con funciones de activación sigmoideas en las capas ocultas y de salida, cuyos resultados indican que las ANN ofrecen una vía prometedora si la medida de rendimiento es el porcentaje de préstamos incobrables clasificados correctamente, pero si la medida de desempeño es el porcentaje de préstamos buenos y malos clasificados correctamente, los modelos de regresión logística son comparables a estos.

2.3. Manejo de conjuntos desequilibrados en ML.

La forma correcta de validar los resultados en escenarios desequilibrados es realizar el procedimiento en los conjuntos de entrenamiento en cada iteración de un procedimiento de validación cruzada, ya que realizar el procedimiento de validación cruzada después del sobremuestreo conduce a resultados sobreoptimistas (Senoae et al., 2018).

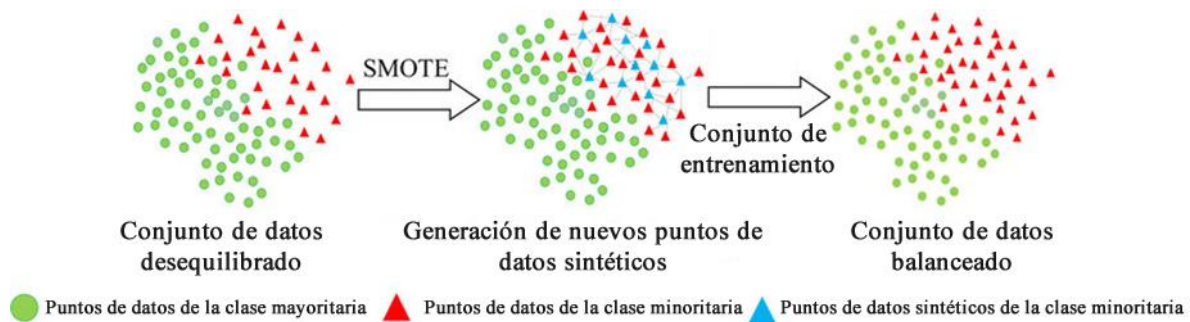


Figura 4. Funcionamiento de la técnica SMOTE.

Fuente: traducido por el autor de Aldraimli et al. (2020).

La técnica SMOTE (Synthetic Minority Oversampling Technique) es una de las técnicas de sobremuestreo más populares, que funciona encontrando un grupo de k vecinos más cercanos en el espacio de características, luego encuentra aleatoriamente un punto dentro de este grupo, y finalmente usa el promedio ponderado para falsificar el nuevo punto de datos (Jiang et al., 2019). Para encontrar la distancia entre los puntos cuando se tienen atributos nominales y continuos Wilson y Martínez (1997) recomiendan utilizar la función HVDM (Heterogeneous Value Difference Metric). En Python se utiliza el paquete SMOTE-NC de la biblioteca imblearn para crear datos sintéticos para variables categóricas y continuas (Aguilar, 2019). En la Figura 4 se ejemplifica el proceso explicado.

2.4. Evaluación de modelos de clasificación binaria en ML.

La matriz de confusión o tabla de contingencia, mide el rendimiento para un problema de clasificación donde la salida puede ser de dos o más clases; cuando se trata de decisión binaria, el clasificador etiqueta ejemplos como positivos o negativos en una tabla (ver Tabla 1) de cuatro combinaciones diferentes de valores predichos y reales: verdaderos positivos (TP), falsos positivos (FP) llamados error de Tipo I, verdaderos negativos (TN) y falsos negativos (FN) llamados error de Tipo II, a partir de la cual se mide la sensibilidad o recuperación (*Recall*), precisión (*Precision*), especificidad, exactitud (*Accuracy*) y lo más importante, las curvas AUC-ROC (*Area Under the Curve – Receiver Operating Characteristic*), que muestran cómo los ejemplos positivos clasificados correctamente varían con la cantidad de ejemplos negativos clasificados incorrectamente; además de las curvas *Precision-Recall* (PR) como alternativa a las ROC para tareas con un gran sesgo en la distribución de clases (Davis y Goadrich, 2006; Narkhede, 2018).

Tabla 1. Matriz de confusión y sus métricas.

		Clases predichas		
		Positivo	Negativo	
Clases observadas	Positivo	Verdadero Positivo (TP)	Falso Negativo (FN) Error de Tipo II	Sensibilidad: $\frac{TP}{TP + FN}$
	Negativo	Falso Positivo (FP) Error de Tipo I	Verdadero Negativo (TN)	Especificidad: $\frac{TN}{TN + FP}$
		Precisión: $\frac{TP}{TP + FP}$	Predicción de negativos: $\frac{TN}{TN + FN}$	Exactitud: $\frac{TP + TN}{TP + TN + FP + FN}$

Fuente: elaboración propia en base a Davis y Goadrich (2006) y Narkhede (2018).

Según Nanni y Lumini (2009) la precisión y los errores de Tipo I y II no son parámetros confiables, puesto que no consideran los puntajes de los clasificadores, por lo que consideran el mejor indicador para comparar métodos las áreas bajo la curva ROC como se sugiere en la literatura sobre aprendizaje automático por Fawcett (2004). Las autoras mejoran los resultados de Tsai y Wu (2008) y en el análisis de sus análisis experimentales concluyen que el mejor clasificador independiente es el MLP.

La puntuación F1 ($F1-score$) es otra métrica importante que combina *Precision* y *Recall*, por definición es la media armónica entre estas, permitiendo medir ambas al mismo tiempo, ya que es difícil comparar dos modelos con baja precisión y alta sensibilidad y viceversa, dado que cuando se intenta aumentar una, la otra disminuye, lo cual se denomina compensación de Precisión/Recuperación (*Precision/Recall trade-off*), y con esta medida se capturan ambas tendencias en un solo valor (Lipton et al., 2014; Narkhede, 2018; Lendave, 2021).

La idea detrás de la compensación es que al cambiar el *threshold* (umbral) para determinar si una clase es positiva o negativa, se inclinará la balanza, y el clasificador funciona de manera distinta para diferentes valores de umbral, lo que significa que la predicción se puede cambiar estableciendo el valor de umbral (Lendave, 2021).

Scikit-Learn (biblioteca de Python) no permite establecer un umbral directamente, pero da acceso a las puntuaciones de decisión que utiliza para hacer predicciones; si el puntaje de decisión es mayor o igual que el valor del umbral, predice la clase positiva; la mayoría de clasificadores usan un umbral igual a 0 (Upadhyay, 2020).

3. METODOLOGÍA.

El enfoque del estudio es de naturaleza cuantitativa, alcance explicativo y de tipo experimental. El conjunto de datos proviene de la base de datos interna de un banco hondureño, y fue obtenido mediante consulta electrónica del reporte de Central de Información Crediticia (CIC) que se entrega al ente regulador de dicho país, llamado Comisión Nacional de Banca y Seguros (CNBS) sobre los préstamos otorgados y que aún no han sido cancelados, cuyos saldos van desde el año 2000 hasta diciembre de 2020.

El conjunto de datos se compone de 51,696 registros y 40 campos que representan las características disponibles para explicar la variable respuesta dicotómica *EnMora*. A continuación, se describe en la Tabla 2, el diccionario de datos basado en la norma local definida en el Manual de Capturador de Crédito de la CNBS (2019):

Tabla 2. Estructura de la tabla de datos de la cartera crediticia.

No.	Variable	Descripción	Tipo de dato	Formato	Escala	Categorías / Valores
1	CODZONA	Se recodificó el campo ZONA con números.	Cualitativo.	Entero.	Nominal, politómica.	1: Occidente, 2: Norte, 3: Centro.
2	CODSUC	Código de sucursal del banco a la que pertenece el préstamo.	Cualitativo.	Entero.	Nominal, politómica.	1 – 9 según la sucursal del banco.
3	CODOFI	Código de la oficina donde fue otorgado el préstamo.	Cualitativo.	Entero.	Nominal, politómica.	Códigos de 3 dígitos para identificar 90 oficinas.
4	TIPOID	Tipo de identificación con el cual se reporta al deudor.	Cualitativo.	Entero.	Nominal, politómica.	7 tipos de identidad, según tabla 2 manual CNBS, pág. 89.
5	TIPPER	Código del tipo de persona del deudor, dependiendo si es persona natural o jurídica.	Cualitativo.	Entero.	Nominal, politómica.	16 tipos de personas, tabla 3 manual CNBS, pág. 89.
6	ESTAOP	Estado de la operación en seguimiento a normas contables.	Cualitativo.	Texto.	Nominal, politómica.	EJE: Ejecución, MOR: Mora, VEN: Vencido, VIG: Vigente.
7	CODESTAOP	Se recodificó ESTAOP con números.	Cualitativo.	Entero.	Nominal, politómica.	1: VIG, 2: MOR, 3: VEN, 4: EJE.
8	FECOTO	Fecha de otorgamiento.	Cuantitativo.	Fecha.	Discreta, de intervalo.	N/A es secuencial.
9	FECVEN	Fecha de vencimiento.	Cuantitativo.	Fecha.	Discreta, de intervalo.	N/A es secuencial.
10	DURACION	Campo calculado como la diferencia en años entre FECVEN y FECOTO.	Cuantitativo.	Entero.	Discreta, de razón.	N/A es de conteo.
11	SALVIG	Saldo vigente.	Cuantitativo.	Decimal.	Continua, de razón.	N/A es continua.
12	MONOTO	Monto otorgado en el préstamo.	Cuantitativo.	Decimal.	Continua, de razón.	N/A es continua.
13	VALGAR	Valor de la garantía.	Cuantitativo.	Decimal.	Continua, de razón.	N/A es continua.
14	NODIAA	Número de días de atraso.	Cuantitativo.	Entero.	Discreta, de razón.	N/A es de conteo.
15	TIPGAR	Tipo de garantía que presenta el préstamo según su naturaleza.	Cualitativo.	Entero.	Nominal, politómica.	16 tipos en la tabla 9 manual CNBS, pág. 117.
16	TIPOPE	Código del tipo de operación a la que corresponde el crédito.	Cualitativo.	Entero.	Nominal, politómica.	40 tipos en la tabla 10 manual CNBS, págs. 117-118.
17	TIPMON	Moneda en que se otorgó el crédito.	Cualitativo.	Entero.	Nominal, dicotómica.	1: Lps, 2: Dls.

18	TIPCRE	Tipo de crédito según normas para clasificación de cartera vigentes.	Cualitativo.	Entero.	Nominal, politómica.	1: Comercial, 2: Consumo, 3: Vivienda, 4: Agropecuario.
19	ORIFON	Origen de los fondos que el banco usa para colocar los créditos.	Cualitativo.	Entero.	Nominal, politómica.	10 tipos, tabla 16, manual CNBS, pág. 123.
20	DESCRE	Destino del crédito.	Cualitativo.	Entero.	Nominal, politómica.	Varios, tabla 22, manual CNBS, págs. 125-130.
21	CREDES	Crédito especial, identifica características particulares del crédito.	Cualitativo.	Carácter.	Nominal, politómica.	Varios, tabla 15, manual CNBS, págs. 122-123.
22	CATCREDES	Se codificó CREDES con números.	Cualitativo.	Entero.	Nominal, politómica.	
23	REFINA	Identifica si el crédito fue refinanciado, readecuado o renovado.	Cualitativo.	Carácter.	Nominal, politómica.	4 tipos, tabla 17, manual CNBS, pág. 123.
24	CATREFINA	Se recodificó REFINA con números.	Cualitativo.	Entero.	Nominal, politómica.	
25	FORPAG	Forma de pago del capital del crédito.	Cualitativo.	Texto.	Nominal, politómica.	28 tipos, tabla 14, manual CNBS, pág. 121.
26	CATFORPAG	Se codificó FORPAG con números.	Cualitativo.	Entero.	Nominal, politómica.	
27	TASINT	Tasa de interés anual nominal.	Cuantitativo.	Decimal.	Continúa, de razón.	N/A es continua.
28	COATPO	Costo Anual Total Porcentual: tasa que incluye el total de costos y gastos inherentes a los créditos, es la tasa de interés anual efectiva.	Cuantitativo.	Decimal.	Continúa, de razón.	N/A es continua.
29	MTOCTA	Monto de la cuota fija nivelada a pagar en el período para la amortización de la deuda.	Cuantitativo.	Decimal.	Continúa, de razón.	N/A es continua.
30	UBICAC	Ubicación geográfica por departamento.	Cualitativo.	Entero.	Nominal, politómica.	18 departamentos, tabla 18 manual CNBS, pág. 124.
31	REGION	Región Urbana o Rural.	Cualitativo.	Carácter.	Nominal, dicotómica.	R: Rural, U: Urbana.
32	CATREGION	Se recodificó REGION con números.	Cualitativo.	Entero.	Nominal, dicotómica.	0: R, 1: U.
33	CATEGF	Categoría asignada a las operaciones del deudor en cumplimiento a la regulación.	Cualitativo.	Texto.	Nominal, politómica.	52 categorías de la tabla 12 del manual CNBS, págs. 119-120.
34	CATRGO	Categoría de Riesgo asignada según la clasificación CATEGF del crédito.	Cualitativo.	Entero.	Ordinal, politómica.	5 niveles, tabla 25 manual CNBS, pág. 132.
35	TPOCLI	Tipo de cliente, según su tamaño y naturaleza del crédito concedido.	Cualitativo.	Entero.	Nominal, politómica.	Grupos de la tabla 12 del manual CNBS, págs. 119-120.
36	MANC	Si el préstamo es individual o a nombre de dos o más clientes.	Cualitativo.	Entero.	Nominal, dicotómica.	0: No, 1: Sí.
37	GENCOD	Género del cliente, se recodificó a número, originalmente estaba como: 'M', 'F' y 'Z' para tipos de personas no naturales.	Cualitativo.	Entero.	Nominal, dicotómica.	0: F, 1: M. Tabla 7 manual CNBS, pág. 97.
38	SALARIO	Registra el ingreso declarado por el cliente en el préstamo.	Cuantitativo.	Decimal.	Continúa, de razón.	N/A es continua.
39	EDAD	Registra la edad del cliente en años.	Cuantitativo.	Entero.	Discreta, de razón.	N/A es de conteo.
40	LABTIE	Tiempo de laborar del cliente en años.	Cuantitativo.	Entero.	Discreta, de razón.	N/A es de conteo.

Fuente: elaboración propia en base a CNBS (2019).

3.1. Técnicas y herramientas.

La consulta de la base de datos fue exportada a un archivo plano e importado a Excel para su limpieza, preparación, filtrado y análisis preliminar exploratorio y estadística descriptiva de los datos. El diseño de los modelos MLP y Logit, su entrenamiento, experimentos y visualización de resultados fue realizado en el lenguaje de programación Python 3.7 empleando el *framework* Anaconda y como editor para los algoritmos JupyterLab 1.0.2. A nivel general, para el análisis y el procesamiento de los datos se utilizaron los paquetes numpy, pandas, matplotlib, imblearn y statsmodels, y para el proceso de ML se usaron las bibliotecas de IA: Keras (la cual utiliza Tensorflow como *backend*) y Scikit-Learn.

3.2. Población y muestra.

La cartera crediticia se delimitó únicamente a los préstamos concedidos a personas naturales, que representan un 73.81% de la cartera total, quedando establecida la población en 38,156 unidades muestrales, las cuales serán analizadas en su totalidad, por lo que el tipo de muestra del estudio será censal. La cartera de préstamos de personas naturales es bastante saludable, ya que solo el 10.38% de estos se encuentra en mora (3,959 préstamos), los restantes 34,197 que representan un 89.62% no están en mora, lo cual implica que las categorías de la variable dependiente no están balanceadas, por lo que será necesario aplicar técnicas estadísticas de balanceo de datos para trabajar con conjuntos desequilibrados.

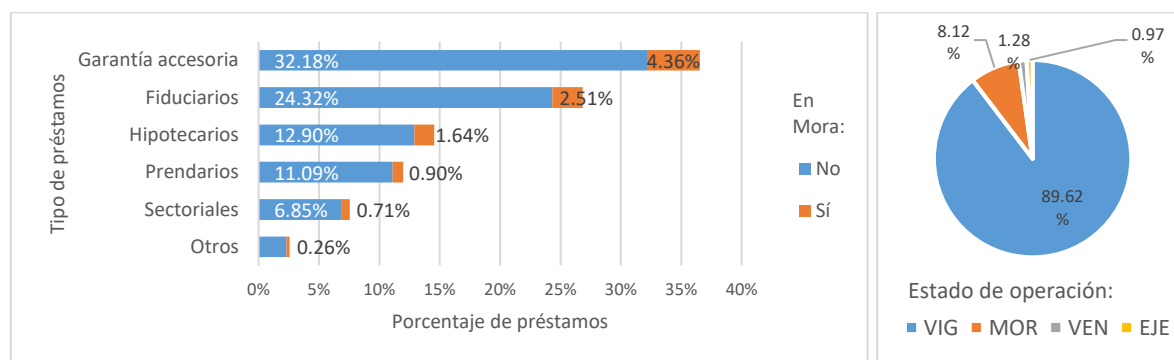


Figura 5. Proporción de la mora según el tipo de préstamo y el estado de operación.

Fuente: elaboración propia del autor.

En cuanto al tipo de préstamos, la mayor parte de los que caen en mora son los concedidos con garantía accesoria, que también es el tipo de préstamo más común dentro de los clientes que son personas naturales, seguidos de los fiduciarios y de los

hipotecarios, estos tres juntos representan 8.31% como se puede ver en la Figura 5, siendo estos grupos en los que se concentra la mayor parte de la mora.

3.3. Identificación y selección de variables.

La variable dependiente para el modelo será de tipo binaria y llamada *EnMora*, la cual indicará si el préstamo está en mora con el valor 1, o no lo está con el valor 0. Entre las 40 características disponibles de la Tabla 1, se preseleccionaron 30 variables independientes por considerarlas adecuadas para predecir el impago en base a las características propias tanto del préstamo, como del cliente. Dentro de estas, 11 eran de naturaleza cuantitativa y 19 eran cualitativas, analizándolas mediante una matriz de coeficientes r de Pearson, para determinar las que presentaban mayor correlación con la variable respuesta.

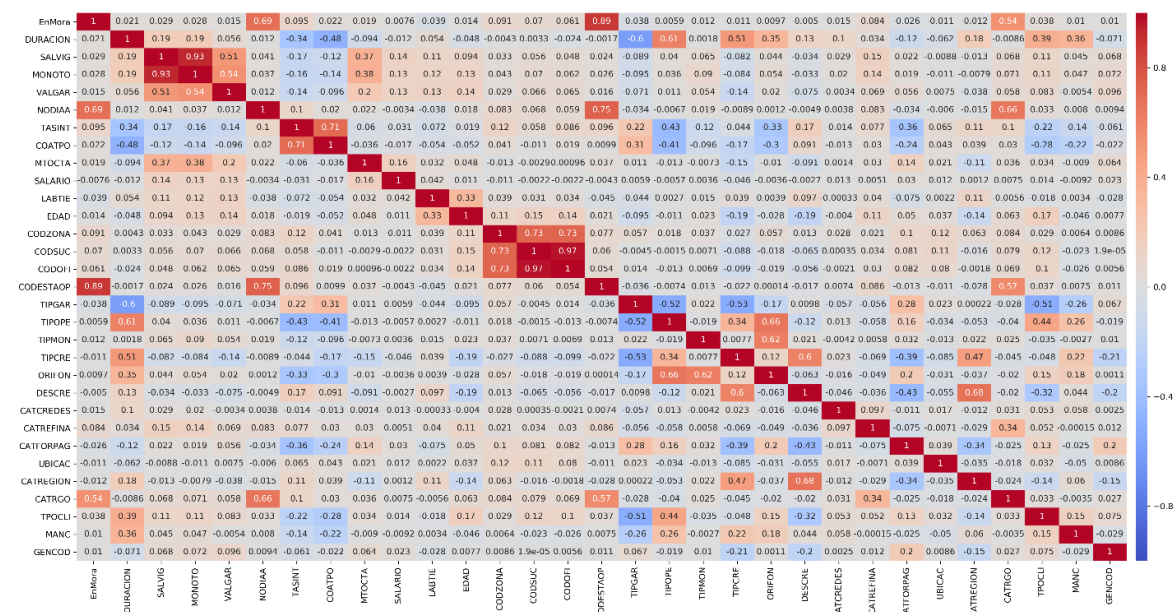


Figura 6. Mapa de calor de la matriz de correlación de las variables de interés para los modelos.

Fuente: elaboración propia del autor.

Las 10 variables que se descartaron de la Tabla 2 fue por las siguientes razones: TIPOID (tipo de identificación) y TIPPER (tipo de persona) solo se utilizaron para filtrar dentro de la población a las personas naturales de las personas jurídicas; asimismo, las variables FECOTO (fecha de otorgamiento) y FECVEN (fecha de vencimiento) solo para estimar la variable DURACION, como la diferencia en años entre estas; mientras

que las variables CREDES (crédito especial), REFINA (refinanciado), FORPAG (forma de pago del capital) y REGION (rural o urbana) como eran todas cualitativas de tipo cadena de caracteres, se codificaron en escala numérica nominal en nuevas variables utilizando los mismos nombres, pero anteponiendo el prefijo CAT a estas. Finalmente, la variable CATEGF (categoría de operaciones del deudor) contenía más del 30% de valores nulos, por lo que también fue descartada.

Como se aprecia en la Figura 6, según la intensidad de la relación, cuando es directamente proporcional se acerca a rojo y cuando es inversamente proporcional a azul, y va variando la tonalidad a medida que se acerca o aleja de cero. En el mapa sobresalen muchas relaciones interesantes entre diferentes variables independientes, pero para efectos del estudio solo relevantes las de la primera columna o fila, que es donde se cruzan todas las independientes con la variable dependiente *EnMora*.

La única variable independiente cuantitativa que presenta una relación fuerte con la mora (0.69) fue el número de días de atraso del préstamo (NODIAA); y entre las cualitativas, solo dos presentaron una correlación importante con la mora, siendo la más fuerte es el estado de la operación (ESTAOP), seguida de la categoría de riesgo (CATRGO) con 0.89 y 0.54 respectivamente. La tercera con mayor correlación es la zona geográfica del país donde se ubica la oficina donde se otorgó el préstamo (CODZONA), la cual engloba la sucursal y el código de oficina, pero con 0.091 no es una correlación fuerte, al igual que las restantes que no llegan a 0.1, como para considerar que su efecto será influyente dentro de los modelos.

Llama la atención que variables como la duración del préstamo en años (DURACION), monto otorgado del préstamo (MONOTO), tasa de interés (TASINT), años de laborar del cliente (LABTIE), edad del cliente (EDAD), y su género (GENCOD), entre otras, que parecen ser de importancia para explicar la mora, y que han sido utilizadas en estudios previos, no presenten alta correlación para este caso de estudio. A pesar de ello, se seleccionaron todas las diez características anteriores por considerarse de interés por los analistas de crédito del banco, para realizar diferentes experimentos entre ellas hasta alcanzar los modelos óptimos, dado que están relacionadas directamente con las fórmulas financieras de valor futuro y cuota fija nivelada, así como las relacionadas para categorizar a los clientes.

4. RESULTADOS OBTENIDOS.

En el análisis exploratorio de los datos en base a las características seleccionadas, al analizar la mora tanto de manera discreta (por conteo de préstamos en cada categoría binaria) como de manera continua (en base a la sumatoria de los montos otorgados), en

relación a la categoría de riesgo asignada a los préstamos, las distribuciones porcentuales de la mora para cada categoría difieren un poco entre sí. Dado que la unidad de medida de la variable respuesta en el modelo a construir es de naturaleza discreta dicotómica, se centrará el estudio en el conteo de préstamos.

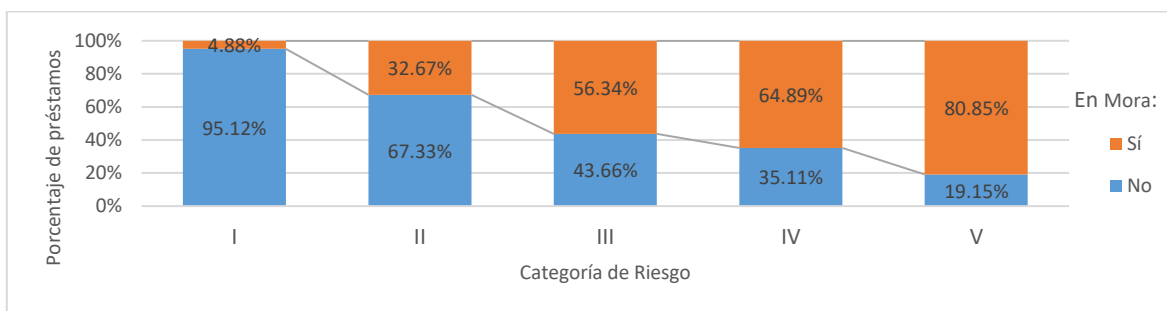


Figura 7. Proporción de préstamos en mora con respecto a la categoría de riesgo.

Fuente: elaboración propia del autor.

Como se puede apreciar en la Figura 7, al visualizar la tendencia que sigue la proporción de préstamos en mora a medida que se va avanzando entre las categorías ordinales de riesgo asignadas, esta va aumentando, mientras que la cantidad de préstamos que no caen en impago va disminuyendo; dicho patrón de comportamiento es el esperado. La clasificación de niveles de riesgo (CATRGO) del deudor según la norma crediticia hondureña es la siguiente: I significa buenos, II especialmente mencionados, III bajo norma, IV dudosa recuperación y V pérdida (CNBS, 2019).

Dicha normativa también establece que el estado vigente (VIG) se refiere a una operación con atraso entre 0 y 89 días, mora (MOR) es un estado de operación cuyo atraso de pagos es mayor o igual a 90 días, vencido (VEN) se considera el estado de operación cuyo plazo estipulado en el contrato o título de crédito ha finalizado, y ejecución judicial (EJE) es cuando la operación ha sido trasladada a instancias judiciales para su recuperación.

En el conjunto de datos, todos los préstamos de personas naturales con estado de operación (ESTAOP) vigente se consideran como no en mora, y cualquier otra categoría de operación abarca los préstamos cuyos clientes han caído en mora; de allí su alta correlación con la variable dependiente, dado que, bajo dicho criterio, en las etiquetas de la variable *EnMora* casi todos los vigentes estarían representados con 0 y los de estado distinto a vigente tomarán el valor de 1.

El 84.41% de la cartera de préstamos al día entra en la categoría I con estado vigente, mientras que 5.21% se reparte entre las categorías II a V del mismo estado de operación, por lo que al abarcar esa gran mayoría (89.62%) se usa una escala logarítmica

base 10 para apreciar mejor las barras de las categorías restantes, dado que las diferencias serían muy grandes en escala normal y no se apreciarían. Los restantes que entran en impago: el 8.12% se categoriza como mora, el 1.28% como vencido y el 0.97% en ejecución, tal y como se visualiza en la Figura 8.

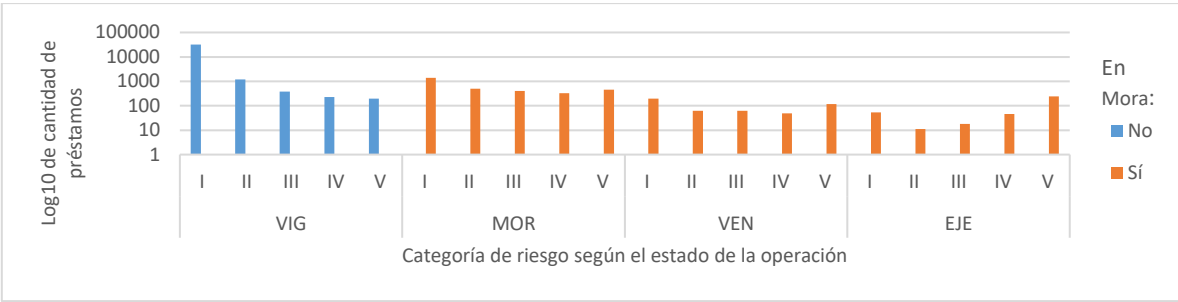


Figura 8. Comportamiento de la mora en préstamos según su estado de operación y categoría de riesgo asignada.

Fuente: elaboración propia del autor.

En cuanto al número de días de atraso, la mora hasta 90 días prácticamente es mínima, pues los préstamos son considerados como vigentes; cuando el estado de la operación es distinto, ya el porcentaje de los préstamos que caen en mora dentro de cada rango de días después de 91 va aumentando, principalmente cuando caen en ejecución judicial, como se aprecia en la Figura 9.

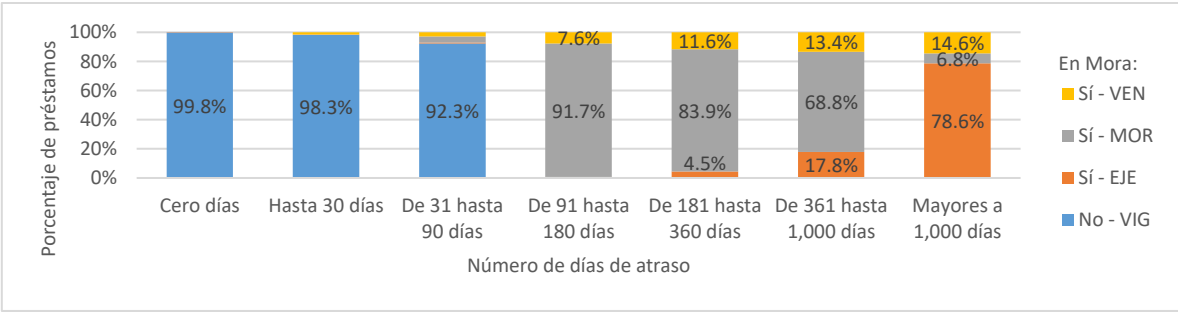


Figura 9. Proporción de los préstamos en mora según los días de atraso y el estado de operación.

Fuente: elaboración propia del autor.

La mayoría de los clientes son hombres, 65% contra solo un 35% de mujeres, los clientes según la normativa pueden clasificarse como: Pequeño Deudor Comercial Hipotecario (PDCH), así como no Hipotecario (PDCnH) y clientes de Consumo y

Vivienda. La mora está concentrada en los PDCH, principalmente en los hombres. El comportamiento general se puede apreciar en la Figura 10.

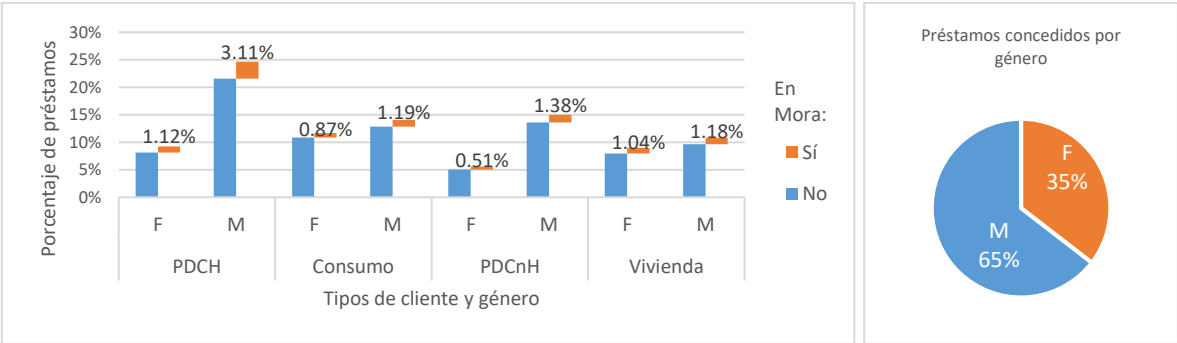


Figura 10. Proporción de la mora según el tipo de cliente y su género.

Fuente: elaboración propia del autor.

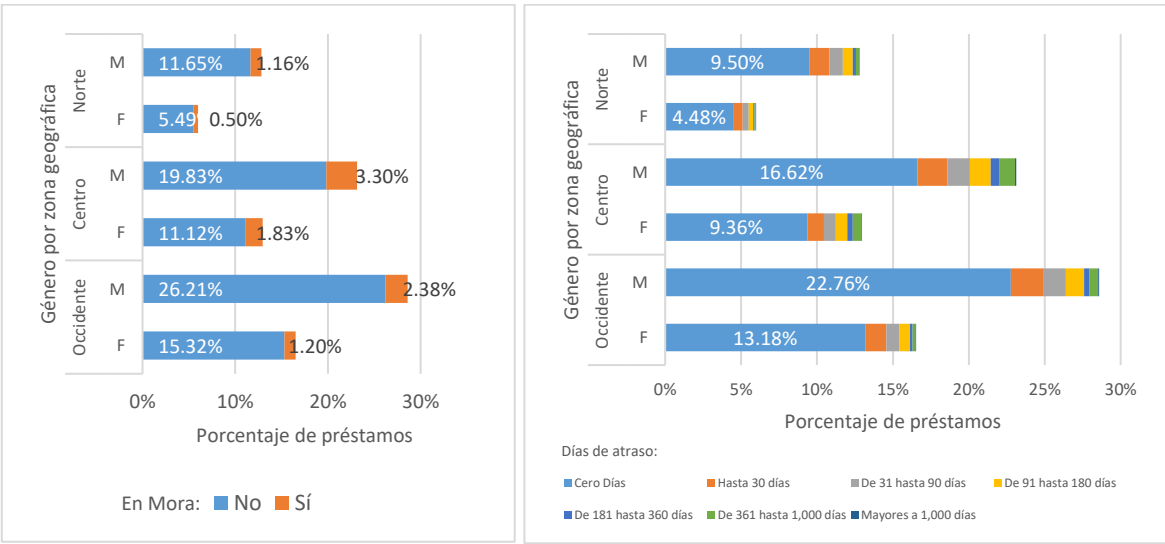


Figura 11. Comportamiento de la mora por zona geográfica, según género y días de atraso.

Fuente: elaboración propia del autor.

La mayor cantidad de préstamos está concentrada geográficamente en la región occidental del país, seguida de la región central (que consolida oficinas de la zona sur y oriental) y por último la zona norte. Al igual que por el tipo de cliente, a nivel geográfico predomina el otorgamiento de préstamos a clientes masculinos. Geográficamente, la cantidad de préstamos en mora están mayormente concentrados en la zona centro

(5.13%), luego en el occidente (3.58%) y el restante 1.66% en el norte. La mayoría de estos están al día con cero días de atraso; la mayoría de clientes que se atrasan hasta 30 o más días en el pago de sus préstamos son hombres ubicados en el centro y occidente del país, siendo la región central la que más días de atrasos presentan sus clientes. Para más información ver Figura 11.

4.1. Resultados de los experimentos con los modelos MLP y Logit.

Se diseñaron varios experimentos hasta llegar a los que entregaban mejores resultados, las variables de entrada seleccionadas para los modelos base MLP y Logit fueron: DURACIÓN, MONOTO, TASINT, LABTIE, EDAD, GENCOD y CODZONA, debido a que constituyen rasgos básicos inherentes a los préstamos, clientes y región geográfica; son conocidas de antemano aún para clientes nuevos que tomarán un préstamo por primera vez y estimar qué tan propensos serían de caer en mora y además, son características declaradas como de especial interés por los analistas de crédito del banco para analizar a sus clientes.

Tabla 2. Diseño de los modelos de red neuronal utilizados en el estudio.

Estructura		Parámetros de los experimentos		
Capa	Nodos	1: Modelo base	2: base + CATRGO	3: base + NODIAA
Entrada	$X =$	7	8	8
Oculto 1 (densa)	20	160	180	180
Oculto 2 (densa)	24	504	504	504
Oculto 3 (dropout)	24	0	0	0
Oculto 4 (densa)	20	500	500	500
Oculto 5 (densa)	24	504	504	504
Salida (densa)	1	25	25	25
Parámetros totales		1,693	1,713	1,713
Parámetros entrenables		1,693	1,713	1,713
Parámetros no entrenables		0	0	0

Fuente: elaboración propia del autor.

La red MLP base consiste en cuatro capas densas ocultas con funciones de activación ReLU y una capa *dropout* intermedia, dicha estructura se determinó mediante prueba y error. El número de nodos de la capa de entrada dependerá del número de variables del experimento y la de salida tendrá un solo nodo con una función de activación sigmoidea, ideal para predecir una variable binaria y es la misma función que utiliza la regresión logística en el modelo Logit de comparación. Por cada experimento se corrieron ambos modelos con la combinación de variables respectivas para tomar las mediciones resultantes y hacer una tabla comparativa con los reportes de clasificación de cada experimento para evaluar su desempeño entre sí.

Las tres variables que resultaron con mayor correlación (CATRGO, NODIAA y ESTAOP) se fueron agregando al modelo base en cada experimento a cada modelo, para analizar la manera en que contribuyen a explicar la mora, tanto individualmente como en combinación. El modelo aprende a identificar de manera perfecta los préstamos en mora cuando conoce el estado de la operación (ESTAOP), lo cual se debe a que todos los que están con estado VIG se encuentran al día, y cualquier estado distinto implicaría que el préstamo ha caído en mora, por lo que no es una característica idónea para predecir; también, al agregar todas las variables restantes al modelo base se obtienen resultados casi perfectos, pero esto porque está incluida dicha característica, por ello se excluye de los modelos finales presentados.

En base a lo anterior, se resumen los diseños finales de tres modelos MLP en la Tabla 2, cada uno de ellos tendrá su equivalente Logit para compararlo. La escala de todas las variables de entrada fue cambiada, en lugar de usar sus valores originales, se escalaron utilizando la función `MinMaxScaler()` de Scikit-Learn, quedando de esa manera todas dentro de un rango entre 0 y 1.

Tabla 3. Reportes de clasificación de los modelos con los subconjuntos de entrenamiento y prueba.

Modelos	MLP						Logit					
	Base		Base + CATRGO		Base + NODIAA		Base		Base + CATRGO		Base + NODIAA	
Experimento	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
TP	23,946	10,251	23,622	10,099	23,887	10,233	23,945	10,251	23,637	10,107	23,946	10,251
TN	0	0	964	444	2,620	1,125	1	1	929	424	1,150	483
FP	2,763	1,196	1,799	752	143	71	2,762	1,195	1,834	772	1,613	713
FN	0	0	324	152	59	18	1	0	309	144	0	0
Exactitud	0.8966	0.8955	0.9205	0.9210	0.9924	0.9922	0.8966	0.8956	0.9198	0.9200	0.9396	0.9377
Sensibilidad	1	1	0.9865	0.9852	0.9975	0.9982	1	1	0.9871	0.9860	1	1
Especificidad	0	0	0.3489	0.3712	0.9482	0.9406	0.0004	0.0008	0.3362	0.3545	0.4162	0.4038
Precisión	0.8966	0.8955	0.9292	0.9307	0.9940	0.9931	0.8966	0.8956	0.9280	0.9290	0.9369	0.9350
Pred. de Neg.	0	0	0.7484	0.7450	0.9780	0.9843	0.5	1	0.7504	0.7465	1	1
F1: 0	0.9455	0.9449	0.9570	0.9572	0.9958	0.9957	0.9455	0.9449	0.9566	0.9566	0.9674	0.9664
F1: 1	0	0	0.4759	0.4955	0.9629	0.9619	0.0007	0.0017	0.4644	0.4807	0.5878	0.5753

Fuente: elaboración propia del autor.

El conjunto de datos se dividió en dos subconjuntos: el de entrenamiento (Train) conformado por el 70% de las unidades muestrales que equivale a 26,709 préstamos, y el de prueba (Test) con el 30% restante equivalente a 11,447. En cuanto al entrenamiento de los modelos MLP, se realizó durante 5 *epochs* y con un *batch size* de 30 para cada uno; tanto la función de pérdida se va reduciendo en cada época como entre todos los modelos

con respecto al base, así como la exactitud va aumentando en la misma manera. Los resultados de los experimentos se presentan en la Tabla 3, los cuales reflejan que ambos modelos bases fueron incapaces de aprender a clasificar los préstamos que podían caer en *default* ($EnMora = 1$) con las características proporcionadas, lo cual se explica por la baja correlación que mantienen con la variable respuesta.

Para reforzar la decisión sobre las características seleccionadas, se realizó una prueba en el modelo base con todas las demás variables que fueron excluidas de la selección final, obteniendo resultados similares, por lo que ninguna de ellas contribuye a mejorar el aprendizaje ni de la red neuronal, ni de la regresión logística base para clasificar el impago de los préstamos.

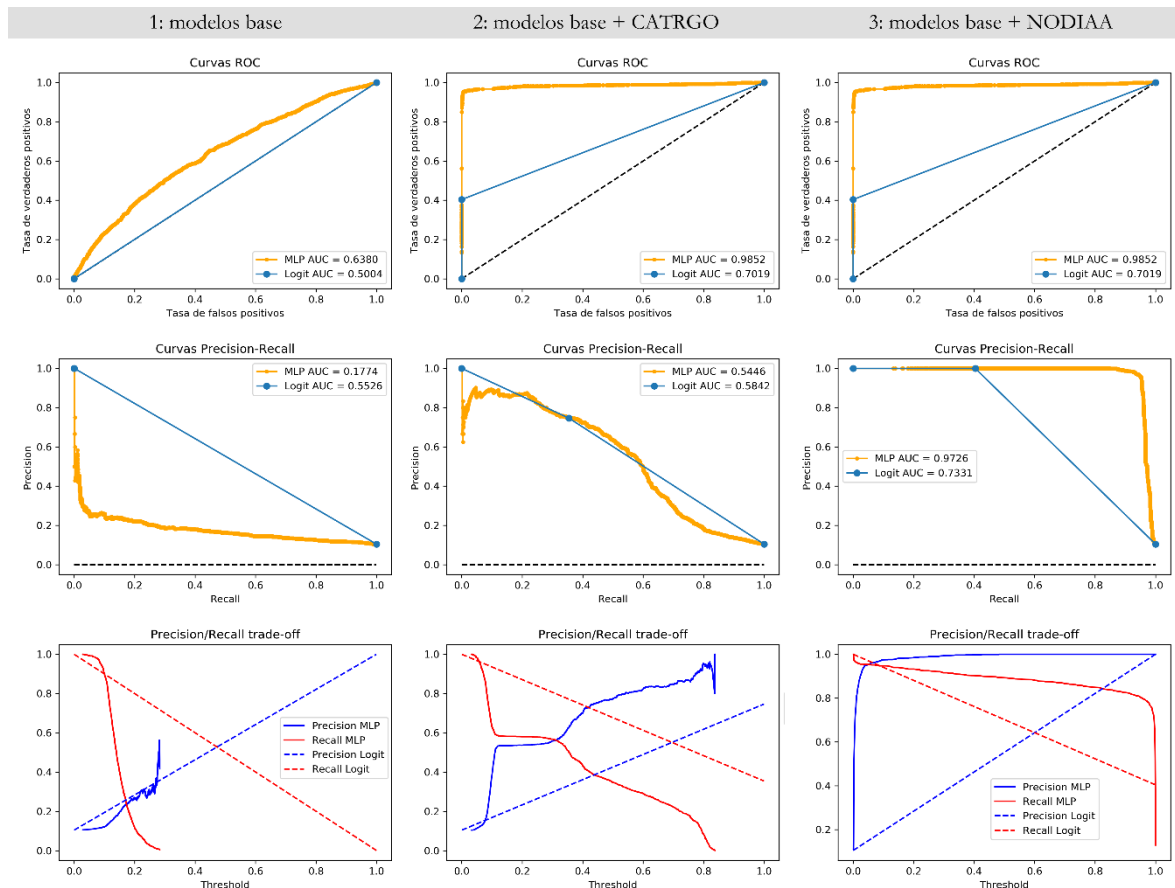


Figura 12. Curvas ROC y PR de los modelos con los subconjuntos de prueba.

Fuente: elaboración propia del autor.

Al agregar la categoría de riesgo del préstamo a los modelos base, estos comienzan a aprender a clasificar los que caen en mora, pero cometen muchos errores de Tipo I y Tipo II, especialmente la regresión logística, situación que se reduce al sustituir esta

variable por el número de días de atraso, las métricas de la matriz de confusión comienzan a demostrar reconocimiento de la clase minoritarias la cual va mejorando entre cada uno de los modelos del mismo tipo, pero el desempeño de los modelos MLP para clasificar correctamente la clase minoritaria resulta superior al de los Logit, también la capacidad predictora de todos va incrementándose entre cada experimento, lo cual se puede apreciar gráficamente en la Figura 12, siendo en ambos casos el modelo 3 MLP el que mejores resultados presenta y se puede representar gráficamente de la siguiente manera:

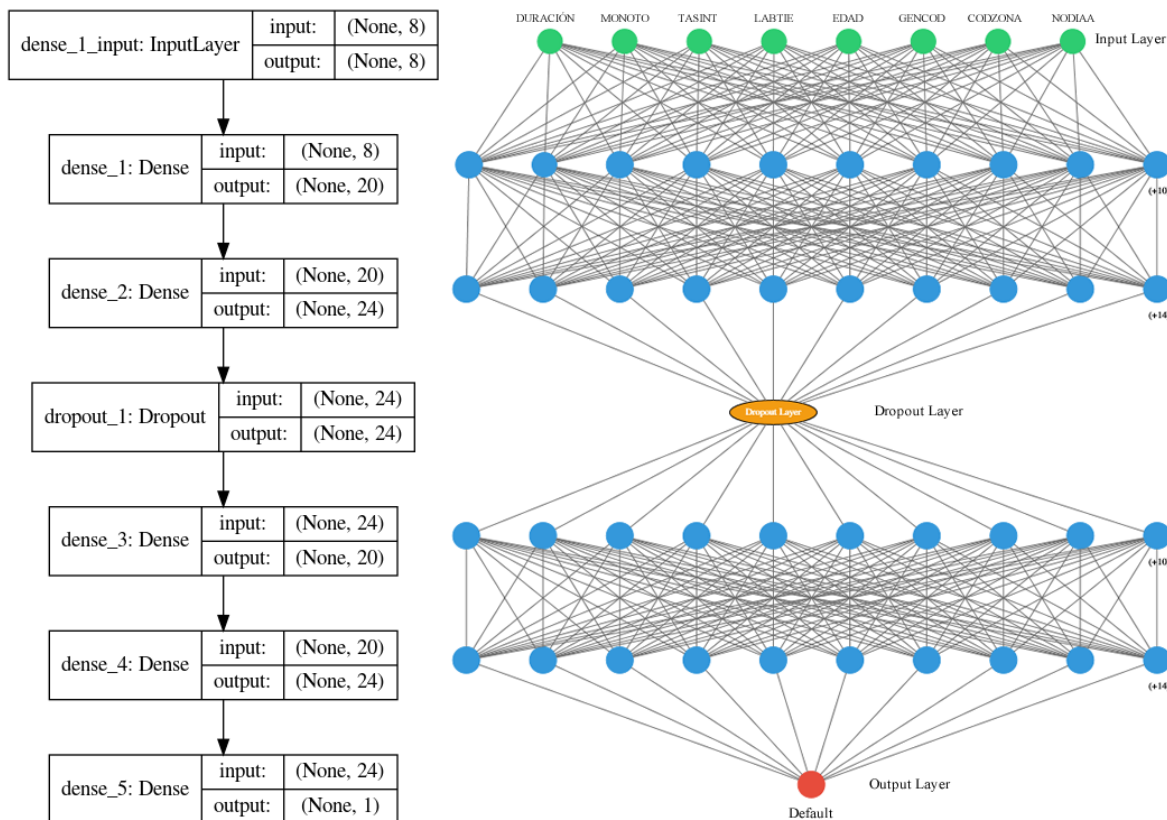


Figura 13. Diagrama del modelo MLP seleccionado.

Fuente: elaboración propia del autor.

Los resultados de los experimentos de los modelos Logit en la Tabla 4, muestran que en el modelo base, todos los coeficientes β de las variables explicativas resultan significativamente distintos de cero; llama la atención que al agregar las variables CATRGO y NODIAA en los subsiguientes experimentos, el p -value del β de la variable MONOTO supera el valor crítico de 0.05 en ambos casos, siendo mayor la probabilidad en el último, lo cual indicaría que al agregar estas variables experimentales al modelo base, el monto del préstamo se volvería innecesario dentro de este para explicar la mora. Además, tanto el AIC como el BIC se van reduciendo sustancialmente entre cada

experimento, lo cual, según criterio de información de Akaike y Bayesiano, se elegiría cada modelo subsiguiente como preferido con respecto al anterior, lo cual se respalda también con el pseudo R^2 que va incrementándose entre cada experimento, siendo el modelo Logit 3 el que mejores resultados presenta dentro de su tipo.

Tabla 4. Resumen de los resultados de los modelos Logit.

1: modelo base	2: modelo base + CATRGO	3: modelo base + NODIAA
Optimization terminated successfully. Current function value: 0.342902 Iterations 7 Results: Modelo base Dataframe: Train	Optimization terminated successfully. Current function value: 0.256020 Iterations 7 Results: Modelo base + CATRGO Dataframe: Train	Optimization terminated successfully. Current function value: 0.037332 Iterations 12 Results: Modelo base + NODIAA Dataframe: Train
Model: Logit Dependent Variable: ENMora Date: 2021-08-12 15:44 No. Observations: 26709 Df Model: 6 Df Residuals: 26702 Converged: 1.0000 No. Iterations: 7.0000	Model: Logit Dependent Variable: ENMora Date: 2021-08-12 14:04 No. Observations: 26709 Df Model: 7 Df Residuals: 26701 Converged: 1.0000 No. Iterations: 7.0000	Model: Logit Dependent Variable: ENMora Date: 2021-08-12 15:54 No. Observations: 26709 Df Model: 7 Df Residuals: 26701 Converged: 1.0000 No. Iterations: 12.0000
Pseudo R-squared: -0.031 AIC: 18331.1178 BIC: 18388.4671 Log-Likelihood: -9158.6 LL-Null: -8883.3 LLR p-value: 1.0000 Scale: 1.0000	Pseudo R-squared: 0.230 AIC: 13692.0561 BIC: 13757.5982 Log-Likelihood: -6838.0 LL-Null: -8883.3 LLR p-value: 0.0000 Scale: 1.0000	Pseudo R-squared: 0.888 AIC: 2018.1869 BIC: 2075.7289 Log-Likelihood: -997.09 LL-Null: -8883.3 LLR p-value: 0.0000 Scale: 1.0000
Coef. Std.Err. z P> z [0.025 0.975]	Coef. Std.Err. z P> z [0.025 0.975]	Coef. Std.Err. z P> z [0.025 0.975]
DURACION -1.0197 0.1050 -9.7141 0.0000 -1.2255 -0.8148	DURACION -0.7542 0.1232 -6.1221 0.0000 -0.9956 -0.5127	DURACION -3.3273 0.3543 -9.3898 0.0000 -4.0218 -2.6327
MONOTO 3.2866 0.6803 4.8312 0.0000 1.9533 4.6199	MONOTO -1.3564 1.2009 -1.1296 0.2587 -3.7101 0.9972	MONOTO 2.3767 3.3726 0.7047 0.4810 -4.2335 8.9869
TASINT -3.9312 0.1333 -29.4980 0.0000 -4.1925 -3.6700	TASINT -4.9894 0.1589 -31.3974 0.0000 -5.3008 -4.6779	TASINT -13.4547 0.4973 -27.0551 0.0000 -14.4294 -12.4800
LABTIE -2.2249 0.2422 -9.1865 0.0000 -2.6996 -1.7502	LABTIE -2.2045 0.2085 -7.6420 0.0000 -2.7699 -1.6391	LABTIE -3.4216 0.7620 -4.4901 0.0000 -4.9132 -1.9281
EDAD -0.7501 0.1153 -6.5077 0.0000 -0.9760 -0.5242	EDAD -1.3624 0.1413 -9.6429 0.0000 -1.6393 -1.0855	EDAD -1.1699 0.3677 -3.1817 0.0015 -1.8996 -0.4492
CODZONA 0.4539 0.0431 10.5336 0.0000 0.3694 0.5383	GENCOD -0.3652 0.0464 -7.8690 0.0000 -0.4561 -0.2742	GENCOD -0.7861 0.1221 -6.4400 0.0000 -1.0254 -0.5469
GENCOD -0.3038 0.0387 -7.8520 0.0000 -0.3796 -0.2280	CODZONA 0.3206 0.0508 6.3068 0.0000 0.2210 0.4203	CODZONA -0.3458 0.1253 -2.7587 0.0058 -0.5914 -0.1001
	CATRGO 5.2807 0.0937 56.3571 0.0000 5.0971 5.4644	NODIAA 360.4526 8.9185 40.4163 0.0000 342.9726 377.9325

Fuente: elaboración propia del autor.

4.2. Optimización de los modelos.

Para balancear el conjunto de datos se aplica un sobremuestreo SMOTE al subconjunto de entrenamiento, quedando cada clase binaria con 23,946 observaciones, obteniendo un *dataframe* de 47,892 registros. Para ello se empleó el paquete SMOTE-NC de la biblioteca imblearn de Python, identificando las variables GENCOD y CODZONA como categóricas.

Tabla 5. Reportes de clasificación de los modelos entrenados con clases balanceadas.

Modelos	MLP						Logit					
Experimento	Base		Base + CATRGO		Base + NODIAA		Base		Base + CATRGO		Base + NODIAA	
Resultados	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
TP	14,496	6,291	22,486	9,596	23,816	10,209	14,676	6,252	22,503	9,612	23,839	10,213
TN	15,343	744	14,129	706	23,231	1,142	13,574	706	14,072	706	22,976	1,114
FP	8,603	452	9,817	490	715	54	10,372	490	9,874	490	970	82
FN	9,450	3,960	1,460	655	130	42	9,270	3,999	1,443	639	107	38
Exactitud	0.6230	0.6146	0.7645	0.9000	0.9824	0.9916	0.5899	0.6078	0.7637	0.9014	0.9775	0.9895
Sensibilidad	0.6054	0.614	0.9390	0.9361	0.9946	0.9959	0.6129	0.6099	0.9397	0.9377	0.9955	0.9963
Especificidad	0.6407	0.622	0.5900	0.5903	0.9701	0.9548	0.5669	0.5903	0.5877	0.5903	0.9595	0.9314
Precisión	0.6276	0.9330	0.6961	0.9514	0.9709	0.9947	0.5859	0.9273	0.6950	0.9515	0.9609	0.9920
Pred. de Neg.	0	0	0.9063	0.5187	0.9944	0.9645	0.5942	0.1501	0.9070	0.5249	0.9954	0.9670
F1: 0	0.6163	0.7404	0.7995	0.9437	0.9826	0.9953	0.5991	0.7358	0.7991	0.9445	0.9779	0.9942
F1: 1	0	0	0.7148	0.5522	0.9821	0.9597	0.5802	0.2393	0.7132	0.5557	0.9771	0.9489

Fuente: elaboración propia del autor.

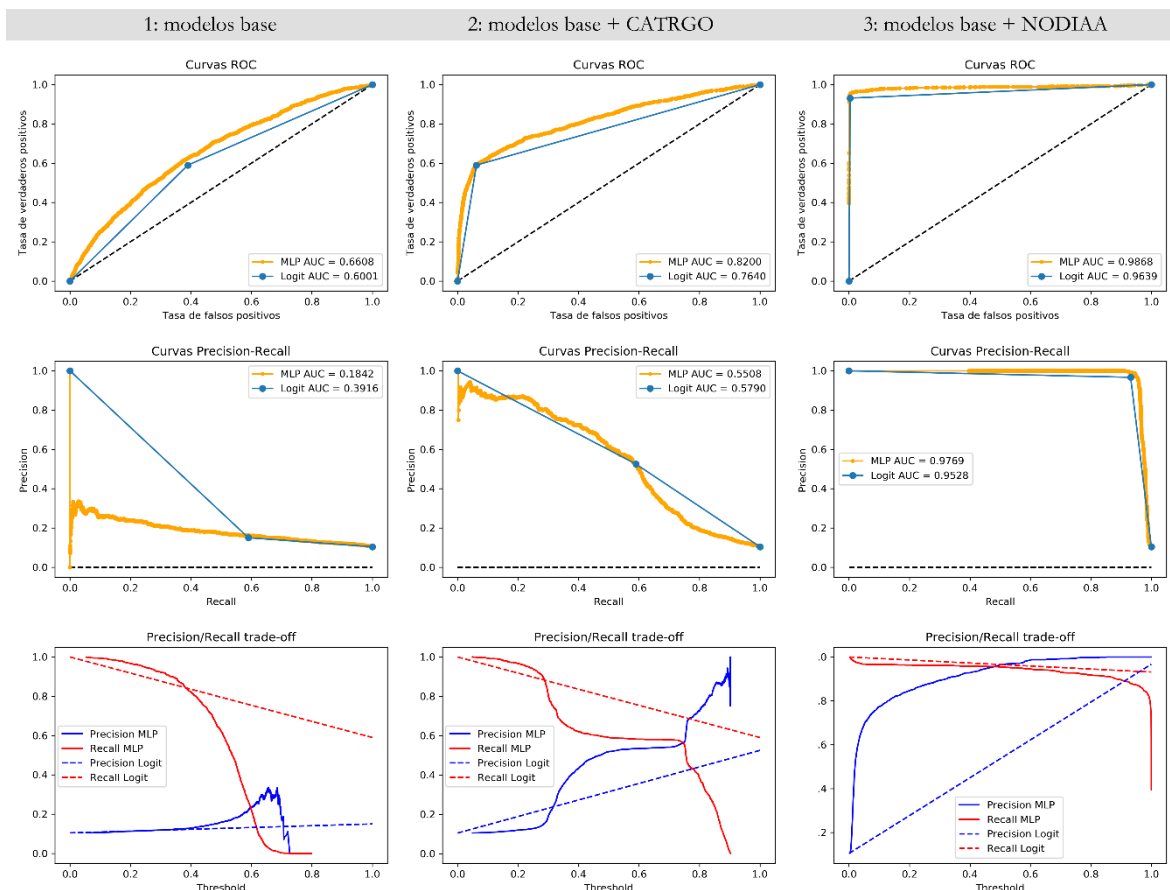


Figura 14. Curvas ROC y PR de los modelos SMOTE los subconjuntos de prueba.

Fuente: elaboración propia del autor.

Al volver a entrenar los modelos con los nuevos datos sintéticos equilibrados en sus clases de salida, y posteriormente pasarle los datos del subconjunto de prueba, se obtuvieron los resultados resumidos en la Tabla 5. El entrenar con clases balanceadas mejora el desempeño de ambos modelos, aunque ligeramente en las redes MLP, pues si bien sus indicadores mejoran, no es tan sustancial como ocurrió en los modelos Logit, donde sí es notable y significativa al acercarlos a los indicadores excepcionales de desempeño alcanzados por los modelos MLP en los experimentos previos, a pesar que no alcanzaron a igualar su desempeño en ninguno de ellos.

En base a dichos resultados graficados en la Figura 14, se elige como modelo preferido el modelo MLP del experimento 3 optimizado con sobremuestreo SMOTE, en el que se combina el modelo base con la variable NODIAA, presentando un área bajo la curva ROC de 98.68% y de 97.69% bajo la curva Precision-Recall y cuya exactitud es 98.24%, precisión del 97.09% y Recall de 99.46% que juntas conforman un F1-Score de 98%, dicho modelo se representa gráficamente igual que el descrito en la Figura 13.

5. DISCUSIÓN Y CONCLUSIONES.

Al evaluar el desempeño de los modelos ANN de tipo MLP contra los Logit, para la clasificación de *default* en préstamos de personas naturales en un banco en Honduras, se descubre que en ambos casos la variable determinante entre todos los experimentos realizados para que los modelos aprendieran a identificar la mora fue el número de días de atraso, siendo el experimento 3 en el que se combinaron las variables base con NODIAA, tanto en los modelos MLP como en los Logit, en el que se obtuvieron los mejores resultados. Sin embargo, el desempeño de las redes neuronales fue superior al de sus equivalentes Logit, principalmente en dicho experimento, dado que aprende mejor a clasificar correctamente la clase minoritaria.

Al aplicar sobremuestreo SMOTE mejora ligeramente el desempeño de las redes MLP, pero no de manera notable, ya que sin necesidad de entrenar con clases balanceadas, presentaron tasas altas de identificación de la clase minoritaria; sin embargo, los modelos Logit sí presentaron una mejora sustancial, especialmente la del último experimento, que casi logra alcanzar en desempeño a su equivalente de red neuronal, por lo que este modelo estadístico tradicional, sí necesita entrenar con clases balanceadas para entregar resultados comparables al de una ANN.

El modelo final elegido, constituye una herramienta de mucha utilidad práctica para los analistas de crédito del banco bajo estudio, puesto que les permitirá conocer de antemano, si un cliente o solicitante que reúna ciertas características ingresadas en las variables de entrada, pueda caer en impago del préstamo que tenga vigente, o que está requiriendo; además de su amplia utilidad metodológica para generar investigación reproducible en otro banco o institución financiera solo adaptando el algoritmo con los nuevos datos de entrada.

6. LIMITACIONES E INVESTIGACIÓN FUTURA.

Al delimitarse a un solo banco, el estudio no entrega un modelo generalizado para todo el sistema financiero del país en cuestión; aunque, la metodología propuesta sí es replicable en cualquier otro banco, cooperativa o empresa financiera del sistema nacional o de cualquier otro país.

De igual manera, el ente regulador, puede diseñar y entrenar un modelo general a partir de la base de datos consolidada que mantiene de todas las entidades supervisadas dentro de su CIC. Asimismo, es recomendable considerar variables macroeconómicas dentro del modelo en investigaciones futuras, para no limitarlo a los datos internos de una institución financiera.

7. REFERENCIAS.

- Aguilar, F. (2019). *SMOTE-NC in ML Categorization Models for Imbalanced Datasets*. Obtenido de Medium Analytics Vidhya: <https://medium.com/analytics-vidhya/smote-nc-in-ml-categorization-models-for-imbalanced-datasets-8adbdcf08c25>
- Aldraimli, M., Soria, D., Parkinson, J., Thomas, E. L., Bell, J. D., Dwek, M. V., & Chaussalet, T. J. (2020). Machine learning prediction of susceptibility to visceral fat associated diseases. *Health and Technology*, 10, 925-944. doi:10.1007/s12553-020-00446-1
- Amidi, A., & Amidi, S. (s.f.). *Deep Learning cheatsheet*. Obtenido de stanford.edu: <https://standord.edu/~shervine/teaching/cs-229/cheatsheet-deep-learning>
- Babs, T. (2018). *The mathematics of Neural Networks*. Obtenido de Medium: <https://medium.com/coinmonks/the-mathematics-of-neural-network-60a112dd3e05>
- Bekhet, H. A., & Eletter, S. F. (2014). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Review of Development Finance*, 4, 20-28. doi:10.1016/j.rdf.2014.03.002
- Brownlee, J. (2016). *Logistic Regression for Machine Learning*. Obtenido de Machine Learning Mastery: <https://machinelearningmastery.com/logistic-regression-for-machine-learning>
- Brownlee, J. (2021). *How to Choose an Activation Function for Deep Learning*. Obtenido de Machine Learning Mastery: <https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/>
- CNBS. (2019). *Manual de Reporte de Datos de Crédito*. Manual, Comisión Nacional de Bancos y Seguros CNBS, Central de Información Crediticia CIC, Tegucigalpa. Obtenido de https://cnbs.gob.hn/wp-content/uploads/2020/02/Ver_2_0_40_MRDC-Capturador-Datos-Credito-SL_NIIF_V_20191209.pdf
- Davis, J., & Goadrich, M. (2006). The Relationship Between Precision-Recall and ROC Curves. *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*. doi:10.1145/1143844.1143874
- Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95, 24-37.
- do Prado, J. W., Carvalho, F. d., Benedicto, G. C., & Lima, A. L. (2019). Analysis of credit risk faced by public companies in Brazil, an approach based on discriminant analysis, logistic regression and artificial neural networks. *Estudios Gerenciales*, 35(153), 347-360. doi:10.18046/j.estger.2019.153.3151

- Golbayani, P., Florescu, I., & Chatterjee, R. (2020). A comparative study of forecasting Corporate Credit Ratings using Neural Networks, Support Vector Machines, and Decision Trees. *arXiv.org*. doi:arXiv:2007.06617v1
- Golbayani, P., Wang, D., & Florescu, I. (2020). Application of Deep Neural Networks to assess corporate Credit Rating. *arXiv.org*. Obtenido de arxiv.org/abs/2003.02334
- Haykin, S. (1999). *Neural Networks A Comprehensive Foundation* (2 ed.). Prentice Hall.
- Haykin, S. (2009). *Neural Networks and Learning Machines* (3 ed.). Pearson.
- Jiang, P., Zhang, J., & Zou, J. (2019). Credit Card Fraud Detection Using Autoencoder Neural Network. *arXiv.org*. doi:arxiv.org/abs/1908.11553
- Keras. (s.f.). *Dropout Layer*. Obtenido de Keras API reference: Regularization layers: https://keras.io/getting_started/faq/
- Keras. (s.f.). *Probabilistic losses*. Obtenido de Keras API Reference: https://keras.io/api/losses/probabilistic_losses/
- Khashman, A. (2010). Neural Networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications*, 37, 6233-6239. doi:10.1016/j.eswa.2010.02.101
- Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *arXiv.org*. doi:arxiv.org/pdf/1412.6980v9.pdf
- Lahsasna, A., Ainon, R. N., & Wah, T. Y. (2010). Credit Scoring Models Using Soft Computing Methods: A Survey. *The International Arab Journal of Information Technology*, 7(2), 115-123. Obtenido de iajit.org/PDF/vol.7,no.2/712.pdf
- Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2), 157-170. doi:10.1016/j.bushor.2019.10.005
- Lendave, V. (2021). *Python Guide to Precision-Recall Tradeoff*. Obtenido de Analytics India Magazine: <https://analyticsindiamag.com/python-guide-to-precision-recall-tradeoff/>
- Lipton, Z., Elkan, C., & Narayanaswamy, B. (2014). Thresholding Classifiers to Maximize F1 Score. *ArXiv.org*. doi:arxiv.org/abs/1402.1892
- Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 36, 3028-3033. doi:10.1016/j.eswa.2008.01.018
- Narkhede, S. (2018). *Understanding Confusion Matrix*. Obtenido de Towards Data Science: <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>

- Obare, D. M., & Muraya, M. M. (2018). Comparison of Accuracy of Support Vector Machine Model and Logistic Regression Model in Predicting Individual Loan Defaults. *American Journal of Applied Mathematics and Statistics*, 6(6), 266-271. doi:10.12691/ajams-6-6-8
- Obare, D. M., Njoroge, G. G., & Muraya, M. M. (2019). Analysis of Individual Loan Defaults Using Logit under Supervised Machine Learning Approach. *Asian Journal of Probability and Statistics*, 3(4), 1-12. doi:10.9734/AJPAS/2019/v3i430100
- Peña, M., & Orellana, J. (2018). Red neuronal para clasificación de riesgo en cooperativas de ahorro y crédito. *Congreso de Ciencia y Tecnología*, 13(1), 121-124. doi:10.24133/ccespe.v13i1.710
- Pérez Ramírez, F. O., & Fernández Castaño, H. (2007). Las Redes Neuronales y la Evaluación del Riesgo de Crédito. *Revista Ingenierías*, 6(10), 77-91. Obtenido de <http://udem.scimago.es/index.php/ingenierias/article/view/225>
- Saavedra García, M. L., & Saavedra García, M. J. (2010). Modelos para medir el riesgo de crédito de la banca. *Cuadernos de Administración*, 23(40), 295-319.
- Salazar Villano, F. E. (2013). Cuantificación del riesgo de incumplimiento en créditos de libre inversión: un ejercicio econométrico para una entidad bancaria del municipio de Popayán, Colombia. *Estudios Gerenciales*, 29, 416-427.
- Senoae Santos, M., Pompeu Soares, J., Henriquez Abreu, P., Araújo, H., & Santos, J. (2018). Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches. *IEEE Computational Intelligence Magazine*, 13(4), 59-76. doi:10.1109/MCI.2018.2866730
- Tiwari, A. K. (2018). Machine Learning Application in Loan Default Prediction. *Jourlan NX - A Multidisciplinary Peer Reviewed Journal*, 4(5), 1-5.
- Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34, 2639-2649. doi:10.1016/j.eswa.2007.05.019
- Upadhyay, A. (2020). *Precision/Recall Tradeoff*. Obtenido de Medium: <https://medium.com/analytics-vidhya/precision-recall-tradeoff-79e892d43134>
- West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, 27, 1131-1152.
- Wilson, D. R., & Martínez, T. R. (1997). Improved Heterogeneous Distance Functions. *Journal of Artificial Intelligence Research*, 6, 1-34.