

## ARTÍCULO ORIGINAL

CREDIT SCORING UNA HERRAMIENTA PARA MINIMIZAR  
EL RIESGO DE CRÉDITO DE LAS INSTITUCIONES  
MICROFINANCIERAS-PERÚCREDIT SCORING A TOOL TO MINIMIZE THE CREDIT RISK OF THE  
MICRO-FINANCIAL INSTITUTIONS-PERU

BALDEMAR QUIROZ CALDERÓN  
Universidad Católica Los Ángeles de Chimbote  
Chimbote, Perú  
ORCID: <https://orcid.org/0000-0002-2286-4606>  
Correo electrónico: [mquirozc@uladech.edu.pe](mailto:mquirozc@uladech.edu.pe)

[Recibido: 24/02/2020 Aceptado: 02/04/2020 Publicado: 07/05/2020]

## RESUMEN

**Objetivo:** Plantear un modelo de Credit Scoring para la cartera de microcréditos de una Caja Municipal de la ciudad de Piura. **Método:** Se aplicó la Regresión Logística Binaria como técnica para plantear un modelo cuya variable respuesta o dependiente es una variable discreta dicotómica. **Resultados:** El tratamiento de la base de datos de la cartera de microcréditos de la Caja Municipal de la ciudad de Piura, mediante el módulo de regresión logística binaria del Software SPSS versión 24, se obtuvo como resultado la probabilidad de default. **Conclusiones:** Se logra obtener un modelo de calificación estadística capaz de predecir correctamente el 96,7% de los créditos de la cartera de la Caja Municipal.

**Palabras clave:** Riesgo; crédito; microfinanzas; tecnología; mercado.

## ABSTRACT

**Objective:** Develop a Credit Scoring model for the microcredit portfolio microcredits from a Municipal Banks in the city of Piura. **Method:** Binary Logistic Regression was applied as a technique to set out a model whose response or dependent variable is a discrete dichotomous variable. **Results:** The treatment of the database of the microcredit portfolio of the Municipal Banks of the city of Piura, using the binary logistic regression module of the SPSS software version 24, was obtained as a result the probability of default. **Conclusions:** it achieved a statistical rating model capable to correctly predict the 96.7% of the credits of portfolio of the municipal bank.

**Keywords:** Risk; Credit; Microfinance; technology; market.

## INTRODUCCIÓN

En el Perú, el éxito y el desarrollo sostenible de las Microfinanzas, ha tenido como base aspectos tan importantes como: el marco regulatorio institucional de apoyo, especialmente normas de información financiera y transparencia, centrales de riesgo, difusión de tasas de interés, tecnologías crediticias apropiadas, la promoción de la transparencia de precios y competencia en el mercado. Sin embargo, este nicho del mercado se hace cada vez más competitivo por efectos naturales de la globalización, expresado en el ingreso de las empresas oligopólicas imponiéndose con nuevos capitales y fundamentalmente con tecnologías actualizadas en la gestión de los riesgos. Lo que pretende este trabajo es proponer un modelo de Credit Scoring para minimizar el riesgo de crédito de la cartera de microcréditos del sector de Microfinanzas-Perú.

Un credit scoring es un sistema de calificación de créditos que intenta automatizar la toma de decisiones en cuanto a conceder o no una determinada operación de riesgo, normalmente un crédito (López, 2019). Un credit rating/credit score (calificación crediticia) es una puntuación que otorgan las agencias de rating a los créditos o deudas de diferentes empresas, Gobiernos o personas, según su calidad crediticia (que mide la probabilidad de que esos créditos sean impagados). La calificación crediticia desempeña un papel fundamental en muchas áreas, como negocios y las finanzas (Fang y Chen, 2019). Si bien las organizaciones de microfinanzas desempeñan un papel importante en las economías en desarrollo, los modelos de apoyo a la toma de decisiones para la calificación crediticia de las microfinanzas no han sido suficientemente cubiertos en la literatura, particularmente para las empresas de microcréditos (Gicić y Subasi, 2019).

También se expone un modelo de calificación crediticia utilizando regresión logística y análisis discriminante multivariante aplicado a 1500 préstamos individuales de IMF marroquíes. Los resultados mostraron la importancia de tener una muestra más grande, un historial lo suficientemente profundo sobre el comportamiento del cliente y también más información sobre las variables relacionadas con la actividad del cliente y su rendimiento para predecir mejor el incumplimiento (Bennouna y Tkiouat, 2019). De igual modo, se presenta el modelo de calificación crediticia aplicado por una institución de microfinanzas en Bosnia y Herzegovina, así como mostrar cómo se identificaron los atributos más relevantes para su implementación (Nalić y Švraka, 2018).

Otro modelo de calificación crediticia proporciona evidencia de una gran institución de microfinanzas (IMF) en India, y hemos aplicado tanto el método de calificación crediticia como el método de red neuronal (NN) y com-

paramos los resultados (Viswanathan y Shanthi, 2017). Los datos no tradicionales pueden usarse para construir algoritmos que puedan identificar a los buenos prestatarios como en la banca tradicional (Ruizy Gama, 2017). El estudio tiene como finalidad proponer un modelo estadístico de Credit Scoring para la cartera de microcréditos del sector de Microfinanzas-Perú. Permitirá mejorar la toma de decisiones para el analista de crédito y reducir el indicador de morosidad de las Microfinancieras. Se observa que estos modelos son utilizados por la Banca comercial. Sin embargo, son limitados los estudios y su aplicación a nivel de las Microfinanzas. La propuesta es elaborar el modelo Credit Scoring para demostrar que si es útil en el tiempo y luego se desarrollará una estrategia que asegure la sostenibilidad del proyecto a largo plazo, bien como un programa lo cual se buscará patentar o como empresa pública o privada.

## MATERIALES Y MÉTODOS

Como metodología se aplicó la Regresión Logística Binaria como técnica para plantear un modelo cuya variable respuesta o dependiente es una variable discreta dicotómica, con un valor de cero (0) cuando el cliente paga (no default) y uno (1) cuando no paga (default) y además  $x_1, x_2, \dots, x_k$ , variables independientes o predictores. La población, 997 clientes que forman parte de la cartera de microcréditos de una Caja Municipal de la ciudad de Piura. Las variables de interés para el estudio fueron: Número de Cuotas Vencidas, Número de Días de Morosidad, Número de Cuotas del Crédito (log), Tasa de Crédito, Edad, Cuentas Pagadas, Antigüedad del Cliente, Número de Créditos en la Institución, Variación Anual del IPC., Sector Económico, Estado Civil, Sexo. Se utilizó el Software SPSS versión 24, el modelo de calificación estadística capaz de predecir correctamente la probabilidad de *default*; esta información sirvió para el diseño del Software de aplicación. Se utilizó la R2 de Cox y Snell y Nagelkerke, la CURVA ROC. (Receiver Operating Characteristic).

El modelo de regresión logística puede formularse como:

$$\text{Log} \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1)$$

Donde:

$p$ : La probabilidad de ocurrencia del evento de interés, en este caso impago o *default*.

$$p = P(\text{default}) = P(\text{impago}) = P(\text{mora}) = P(\text{Incumplimiento})$$

$1-p$ : Es el complemento de la probabilidad de ocurrencia del evento de interés, en este caso es la probabilidad que el cliente pague o *no default*.

$\frac{p}{1-p}$ : Se le llama **Odds**. Se define como la proporción del

número esperado de veces que ocurra un evento y el número esperado de veces que no ocurra.

Dado el valor de las variables independientes, la probabilidad de *default*, puede ser calculada directamente de la siguiente forma (Wiginton 1980):

$$p = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}$$

## RESULTADOS

### Evaluación de la significancia conjunta

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_k$$

$$H_1: \beta_i \neq \beta_j \text{ para todo } i \neq j$$

La tabla 1 muestra el valor de significancia conjunta del modelo  $0,000 < 0,05$ ; por consiguiente se rechaza la hipótesis  $H_0$ ; entonces, existirá por lo menos un  $\beta_i \neq 0$ . Es decir, existe por lo menos una variable independiente que explique el comportamiento de la variable dependiente (*default*).

### Evaluación de la significancia individual

En la tabla 2 se observan las variables especificadas en el paso 1:

Número de Cuotas Vencidas, Número de Días de Morosidad, Número de Cuotas del Crédito (log), Tasa de Crédito, Edad, Cuotas Pagadas, Antigüedad del Cliente, Número de Créditos en la Institución, Variación Anual del IPC., Sector Económico Dico 0,1, Estado Civil Dico 0,1, Sexo.

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

$$\text{Test de Wald} = \left(\frac{\hat{\beta}}{S_{\hat{\beta}}}\right)^2 \sim \chi_1^2$$

### Evaluación del ajuste del modelo

En la tabla 3 se observa que la estimación <sup>a</sup> ha terminado en el número de iteración 8 porque las estimaciones de parámetro han cambiado en menos de ,001.

Para evaluar la bondad del ajuste del modelo de Regresión Logística, deben observarse los indicadores:

Tabla 1  
Pruebas ómnibus de coeficientes de modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	705,855	12	,000
	Bloque	705,855	12	,000
	Modelo	705,855	12	,000

Fuente: Elaboración propia

Tabla 2  
Variables en la ecuación

	B	Error estándar	Wald	gl	Sig.
Número de Cuotas Vencidas	-7,181	4,086	3,088	1	0,079
Número de Días de Morosidad	13,045	1,625	64,419	1	0,000
Número de Cuotas del Crédito	-5,803	3,366	2,972	1	0,085
Tasa de Crédito	-15,253	4,189	13,257	1	0,000
Edad	3,014	3,116	,936	1	0,333
Cuotas Pagadas	-3,584	2,082	2,963	1	0,085
Antigüedad del Cliente	,452	1,295	,122	1	0,727
Número de Créditos en la Institución	,999	1,573	,403	1	0,525
Variación Anual del IPC.	-21,485	12,309	3,047	1	0,081
Sector Económico(1)	,052	,397	,017	1	0,895
Estado Civil 1)	,524	,415	1,595	1	0,207
Sexo(1)	,883	,406	4,725	1	0,030
Constante	16,373	6,307	6,738	1	0,009

Fuente: Elaboración propia

$R^2$  de Cox y Snell; este indicador muestra que el modelo explica el 50,7% de la variabilidad de la variable dependiente (*default*).

$R^2$  de Nagelkerke; este indicador muestra que el modelo explica el 82,9% de la variabilidad de la variable dependiente (*default*).

Tabla 3  
Resumen del modelo  $R^2$  de Cox y Snell y Nagelkerke

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	238,758 <sup>a</sup>	0,507	0,829

Fuente: Elaboración propia

Tabla 4  
Tabla de clasificación

Observado	Pronosticado			Porcentaje correcto
		Calificación del Cliente		
		No Default	Default	
Calificación del Cliente	No Default	807	9	98,9
	Default	24	157	86,7
Porcentaje global				96,7

Fuente: Elaboración propia

- Sensibilidad: Porcentaje de *default* que son clasificados como *default*.

$$\text{Sensibilidad} = \frac{157}{181} = 86,7\%$$

- Especificidad: Porcentaje de *no default* que son clasificados como *no default*.

$$\text{Especificidad} = \frac{807}{816} = 98,9\%$$

Según los resultados de la tabla de clasificación, de los 181 clientes que no cumplen con sus pagos (*default*), el modelo estimó correctamente 157 de ellos; es decir, el 86,7%. Por tanto, el modelo tiene una sensibilidad del 86,7%.

De los 816 clientes que cumplen con sus pagos (*no default*), el modelo estimó correctamente a 807 de ellos; es decir al 98,9%. Por tanto, el modelo tiene una especificidad del 98,9%.

Finalmente, si sumamos 807 (no default) más 157 (default) es igual a 964 clientes que fueron seleccionados correctamente, del total de 997. Por tanto, la probabilidad que se tiene de acertar con este modelo de regresión logística es del 96,7%.

### Curva Roc. (Receiver Operating Characteristic)

El análisis ROC, permite evaluar la precisión de algunos modelos estadísticos, tales como: Regresión Logística, Análisis Discriminante, etc.

Es un gráfico en el que en el eje de las X se representa (*1-Specificidad*) y en el eje de las Y, la *Sensibilidad*.

La curva ROC, cuantifica la capacidad de un indicador, en este caso el *default*, para discriminar entre buenos y malos pagadores. Es decir, mientras mayor sea el área bajo la curva ROC, el modelo de Regresión Logística tiene un mayor poder de discriminación. El resultado arrojado para este modelo fue de 96,6%

### Resumen de procesamiento de casos

La clasificación del cliente positivo (estado real positivo es Default) con N válido igual a 181. De igual modo, el cliente negativo tiene un N válido igual a 816. Los valores más grandes de las variables de resultado de prueba indican una prueba mayor para un estado real positivo.

En la figura 1 se muestran las variables de resultado de prueba: Probabilidad pronosticada con un Área igual a 0,966.

### Interpretación del modelo

Para la interpretación del modelo se tiene que analizar simultáneamente los  $\beta_i$  y sus correspondientes  $\text{Exp}(\beta_i)$ .

Además:

$$\text{Odds Ratio} = OR = \text{Exp}(\beta_i) = e^{\beta_i}$$

Entonces:

- Si  $\beta_i > 0$ , implica que el  $OR > 0$ . (la variable  $X_i$  aumenta la probabilidad de impago).
- Si  $\beta_i < 0$ , implica que el  $OR < 0$ . (la variable  $X_i$  disminuye la probabilidad de impago).
- Si  $\beta_i = 0$ , implica que el  $OR = 1$  (la variable  $X_i$  no ejerce ningún efecto sobre la probabilidad de impago).

$X_1$  = **Número de Cuotas Vencidas**. Por cada unidad de incremento del número de cuotas vencidas, es 0.001 veces menos probable que el cliente caiga en default, frente a no default.

$X_2$  = **Número de Días de Morosidad**. Por cada día de aumento en la Morosidad del cliente, es 462,998.4 veces más probable que el cliente caiga en default, frente a no default.

$X_3$  = **Número de Cuotas del Crédito**. Por cada unidad de incremento del número de cuotas del Crédito, es 0,003 veces menos probable que el cliente caiga en default, frente a no default.

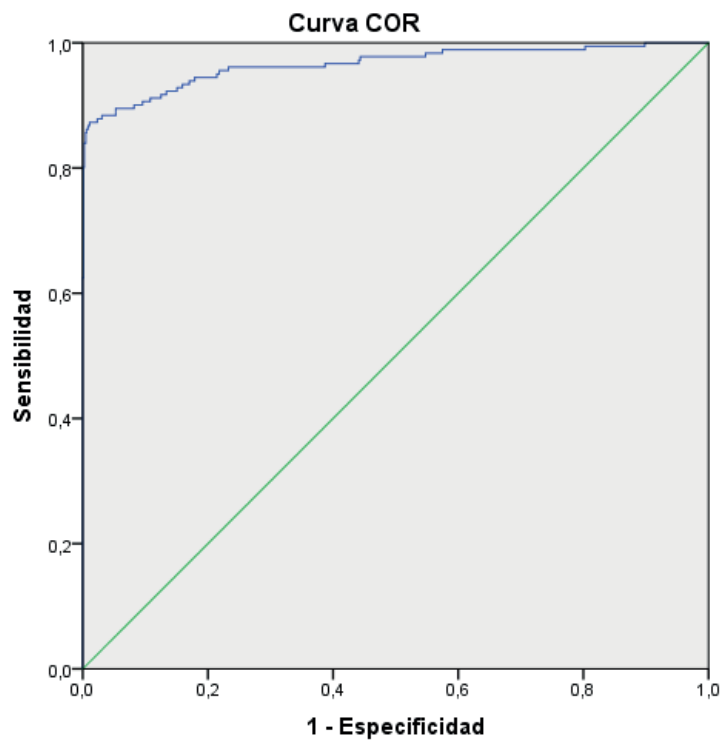


Figura 1. Área bajo la curva de probabilidad pronosticada

Fuente: Elaboración propia

**X<sub>4</sub> = Tasa de Crédito.** Por cada unidad de incremento en la tasa del crédito, es 0,000 veces menos probable que el cliente caiga en default, frente a no default.

**X<sub>5</sub> = Edad.** Por cada año de aumento en la edad del cliente, es 20.4 veces más probable que el cliente caiga en default, frente a no default.

**X<sub>6</sub> = Cuotas Pagadas.** Por cada unidad de incremento en el número de cuotas pagadas del cliente, es 0,028 veces menos probable que el cliente caiga en default, frente a no default.

**X<sub>7</sub> = Antigüedad del Cliente en años.** Por cada año de aumento en la antigüedad del cliente, es 1,571 veces más probable que el cliente caiga en default, frente a no default.

**X<sub>8</sub> = Número de Créditos en la Institución.** Por cada unidad de aumento del número de créditos del cliente, es 2,716 veces más probable que el cliente caiga en default, frente a no default.

**X<sub>9</sub> = Variación Anual del IPC.** Por cada unidad de incremento en el IPC anual, es 0,000 veces menos probable que el cliente caiga en default, frente a no default.

**X<sub>10</sub> = Sector Económico (1).** Que un cliente no pague su deuda (default) es 1.054 veces más probable en todos<sup>1</sup> que los clientes que se dedican al comercio.

**X<sub>11</sub> = Estado Civil (1).** Que un cliente no pague su deuda (default) es 1,689 veces más probable en los solteros<sup>1</sup> que los clientes casados, viudos, divorciados, etc.

**X<sub>12</sub> = Sexo (1).** Que un cliente no pague su deuda (default) es 2.419 veces más probable en los hombres<sup>1</sup> que las mujeres.

Finalmente:

$$p = P(\text{default}) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}} \quad (4)$$

Donde:

$$z = 16,373 - 7,181x_1 + 13,045x_2 - 5,803x_3 - 15,253x_4 + 3,014x_5 - 3,584x_6 + 0,452x_7 + 0,999x_8 - 21,485x_9 + 0,052x_{10} + 0,524x_{11} + 0,883x_{12} \quad (5)$$

Además:

X<sub>1</sub> = Número de Cuotas Vencidas

X<sub>2</sub> = Número de Días de Morosidad

X<sub>3</sub> = Número de Cuotas del Crédito

X<sub>4</sub> = Tasa de Crédito

$X_5$ = Edad

$X_6$ = Cuotas Pagadas

$X_7$ = Antigüedad del Cliente en años

$X_8$ = Número de Créditos en la Institución

$X_9$ = Variación Anual del IPC.

$X_{10}$  = Sector Económico (1)

$X_{11}$  = Estado Civil (1)

$X_{12}$ = Sexo (1)

El tratamiento de la base de datos de la cartera de microcréditos de la Caja Municipal de la ciudad de Piura, mediante el módulo de regresión logística binaria del *Software* SPSS versión 24, se obtuvo como resultado la probabilidad de *default* (expresión 4), conjuntamente con la ecuación de puntuación (expresión 5).

En cuanto a la valoración del modelo y a la vista de los coeficientes del mismo mostrados en la tabla 5, se observa que 05 variables influyen positivamente y 07 influyen negativamente en la probabilidad de que un cliente sea moroso. Cabe destacar, que la variable “Variación Anual del IPC”, correspondiente al grupo de variables macroeconómicas, forma parte de este modelo.

Respecto a la bondad de ajuste, según la tabla 3, se encontró que los  $R^2$  de Cox y Snell y  $R^2$  de Nagelkerke, fueron de

50,7% y 82,9%; respectivamente. Estos indicadores muestran que el modelo explica el 82,9% de la variabilidad de la variable dependiente (*default*).

A su vez, es interesante valorar la capacidad o eficacia predictiva del modelo, lo que se realiza a partir de la tabla 4. La sensibilidad del modelo fue del 86,7%. Así mismo, el porcentaje correcto de clasificación obtenido para este modelo fue del 96,7%, valor que es concordante con área del 0,966 obtenida en la figura 1 de la curva ROC.

## DISCUSIÓN

Una revisión bibliográfica de la literatura sobre modelos para instituciones de Microfinanzas en nuestro país, ha puesto de manifiesto la insuficiencia de trabajos en este campo de estudio, lo que permite afirmar la necesidad de investigaciones futuras que aborden este tema y establezcan propuestas de modelos eficientes. Sin embargo, considerando la revisión de las investigaciones empíricas focalizadas en la evaluación del riesgo de crédito en las instituciones de microfinanzas en Latinoamérica realizado por Seijas-Giménez et al. (2017), el análisis de los modelos de credit scoring permite advertir que pueden ser utilizados en varias dimensiones. No obstante, en general, se observa una preferencia por predecir el riesgo de que los microcréditos incurran en algún tipo de atraso costoso y sobre el cual la IMF tenga control a efectos de mitigarlo. Asimismo, dentro del conjunto de técnicas denominadas credit scoring, la revisión de la literatura

Tabla 5

*Variables en la ecuación*

	B	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
Número de Cuotas Vencidas	-7,181	4,086	3,088	1	0,079	0,001	0,000	2,290
Número de Días de Morosidad	13,045	1,625	64,419	1	0,000	462998,4	19145,2	11196913,0
Número de Cuotas del Crédito	-5,803	3,366	2,972	1	0,085	0,003	0,000	2,214
Tasa de Crédito	-15,253	4,189	13,257	1	0,000	0,000	0,000	,001
Edad	3,014	3,116	0,936	1	0,333	20,4	0,000	9155,0
Cuotas Pagadas	-3,584	2,082	2,963	1	0,085	0,028	0,000	1,644
Antigüedad del Cliente	0,452	1,295	0,122	1	0,727	1,571	0,124	19,892
Número de Créditos en la Institución	0,999	1,573	0,403	1	0,525	2,716	0,124	59,289
Variación Anual del IPC.	-21,485	12,309	3,047	1	0,081	0,000	0,000	14,010
Sector Económico (1)	0,052	,397	0,017	1	0,895	1,054	0,484	2,294
Estado Civil (1)	0,524	,415	1,595	1	0,207	1,689	0,749	3,807
Sexo(1)	0,883	,406	4,725	1	0,030	2,419	1,091	5,365
Constante	16,373	6,307	6,738	1	0,009	12903240,0		

Fuente: Elaboración propia

realizada permite afirmar que existe una preferencia por las técnicas paramétricas en el ámbito de los microcréditos destacando especialmente la regresión logística. Sin embargo, en los últimos años se registra un mayor número de trabajos que utilizan técnicas no paramétricas y que evalúan alternativamente la performance de varias de ellas respecto a una misma base de datos de préstamos o prestatarios. Los resultados muestran que las técnicas no paramétricas logran un mayor poder predictivo del incumplimiento de los clientes de microcréditos. Es por ello que, se ha planteado un modelo de *credit scoring* para la cartera de microcréditos de una Caja Municipal en la que, aplicando la regresión logística binaria, se ha logrado obtener un modelo de calificación estadística capaz de predecir correctamente el 96,7% de los créditos de la cartera.

## REFERENCIAS

- Bennouna, G. y Tkiouat, M. (2019). Scoring in microfinance: credit risk management tool—Case of Morocco. *Procedia computer science*, 148, 522-531.
- Fang, F. y Chen, Y. (2019). A new approach for credit scoring by directly maximizing the Kolmogorov–Smirnov statistic. *Computational Statistics & Data Analysis*, 133, 180-194.
- Gicić, A. y Subasi, A. (2019). *Calificación crediticia para un conjunto de datos de microcrédito utilizando la técnica de sobremuestreo de minorías sintéticas y clasificadoras de conjuntos*. DOI: 10.1111/exsy.12363. Recuperado de: <https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12363>
- López, I. (2019). *Credit scoring*. Diccionario económico. En *expansion.com*. Recuperado de: <https://www.expansion.com/diccionario-economico/credit-scoring.html>
- Nalić, J. y Švraka, A. (2018). Using data mining approaches to build credit scoring model: Case study—Implementation of credit scoring model in microfinance institution. In 2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH), (pp. 1-5). IEEE.
- Ruiz, S. y Gama, J. (2017). Credit Scoring in Microfinance Using Non-traditional Data. In *EPIA Conference on Artificial Intelligence*, (pp. 447-458). Springer, Cham.
- Seijas-Giménez, M., Vivel-Búa, M., Lado-Sestayo, R. y Fernández-López, S. (2017). La evaluación del riesgo de crédito en las instituciones de microfinanzas: Estado del arte. *Compendium: Cuadernos de Economía y Administración*, 4(9), 35-52.
- Viswanathan, P. y Shanthi, S. (2017). Modelling Credit Default in Microfinance—An Indian Case Study. *Journal of Emerging Market Finance*, 16(3), 246-258.

