

UN MODELO DE CREDIT SCORING BASADO EN EL CONOCIMIENTO DE LA APLICACIÓN DE BASILEA II Y SU PAPEL INNOVADOR EN EL SECTOR BANCARIO

Dra. Elena M^a Esteve

1.INTRODUCCIÓN

El objetivo de este trabajo es desarrollar un modelo de Credit Scoring que, además de cumplir las características tradicionales de calificación y clasificación de los clientes, incorpore todos aquellos requisitos necesarios para ser considerado un modelo interno para la valoración de riesgo, como es requerido por el Nuevo Acuerdo de Basilea II [9].

Así, el modelo de Credit Scoring que planteamos, nos debe permitir obtener una clasificación de los clientes en segmentos. Cada uno de estos deberá reunir unos mismos patrones de riesgo, de manera que el modelo construido permita calcular ó bien la pérdida esperada (EL), ó bien, los valores de la probabilidad de impago (PD) y de la pérdida dado impago (LGD), para cada segmento de clientes obtenido.

El modelo propuesto combina un método estadístico y un modelo de red neuronal, concretamente Análisis Discriminante y el Algoritmo de Kohonen.

En la sección 4 de este artículo, se explica como se utilizan los métodos de Análisis Discriminante y el Algoritmo de Kohonen en la elaboración de nuestro modelo de Credit Scoring, detallando los conceptos fundamentales que se emplean, así como la definición de las componentes de riesgo establecidas por Basilea II, y su relación con el modelo de Credit scoring desarrollado. En la sección 5 se resuelve un caso práctico de aplicación del modelo propuesto y se exponen las conclusiones que se derivan de los resultados obtenidos.

2.METODOLOGÍA

En esta sección, presentamos nuestro modelo de Credit Scoring. Primero desarrollamos el modelo mixto teórico, y después realizamos una aplicación empírica del mismo, al mismo tiempo que calculamos las componentes de riesgo propuestas por Basilea II, para que este modelo pueda ser considerado un modelo interno de cálculo de riesgo.

2.1 Análisis discriminante

Las principales aportaciones a nuestro credit scoring del análisis discriminante son:

La aportación fundamental es poder obtener los valores de los coeficientes estandarizados a_{1j}^* , porque nos permiten establecer un orden de incidencia de las características sobre el pago del crédito, tanto es así que a partir de ahora vamos a utilizar el valor de estos coeficientes para denotar el peso de las variables X_i en el pago o impago de los préstamos.

La utilización de estos coeficientes solventa al mismo tiempo, el problema que presentan las redes neuronales, puesto que estas no distinguen las características influyentes en la solvencia del préstamo de las que no lo son.

- La utilización del análisis discriminante no solo nos cuantifica la importancia de las variables, sino que nos cualifica a las mismas, esto es, nos dice si las variables relevantes son p ó $d \leq p$
- Además, el análisis discriminante como técnica de clasificación de datos nos proporciona una clasificación inicial de clientes, y con la elaboración de nuestro modelo queremos aportar que al añadir el método de redes neuronales esta primera clasificación se puede mejorar, por lo tanto nos aporta la posibilidad de comparar los resultados obtenidos solo con análisis discriminante con los conseguidos por nuestro modelo mixto.

Sin embargo, el análisis discriminante no puede aportar una segmentación fundamentada del riesgo, y esta será la aportación principal del modelo de redes neuronales que vemos en la siguiente sección

2.2 Elaboración del Modelo

En la elaboración del modelo propuesto se planteaba la posibilidad de recurrir a valoraciones del cliente ya realizadas, o bien, construir una nueva forma de calificación. Nosotros optamos por la segunda por dos razones, la primera es que esto supone una aportación, creemos que importante, para este trabajo y la segunda razón es que así, partíamos de una valoración que nos servía para la siguiente parte del modelo que consiste en la clasificación del cliente y en el cálculo de la probabilidad de impago.

La obtención de la nota del cliente no ha sido sencilla, en un primer momento nos planteamos aplicar el algoritmo de Kohonen en cada una de las características y hacer clasificaciones de cada una de ellas, de manera que cada clase conseguida en una característica le asignáramos tanto el valor de los coeficientes estandarizados obtenidos en el análisis discriminante como en el algoritmo, pero este método de calificación, además de ser relativamente complicado presentaba problemas con las características que inciden negativamente en el préstamo. Con la misma perspectiva de seguir utilizando los dos métodos para la calificación se elaboró una formulación parecida a la definitiva con el hándicap de que no permitía obtener una función continua de las mismas, lo que nos llevaba a perder muchas posibles propiedades con las que poder trabajar con posterioridad con esta función, y tras múltiples pruebas llegamos a la conclusión de que la más satisfactoria es la que expresamos a continuación

$$U=U(X_1, X_2, \dots, X_d) = (1 / \sum_{j=1}^d |a_{1j}^*|) \sum_{j=1}^d a_{1j}^* (I_{(0,\infty)}(a_{1j}^*) - X_j)$$

donde

$$I_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Para obtener el valor de U para todos y cada uno de los clientes cuyas características ya conocemos y sabemos como han respondido ante el préstamo concedido, lo primero que tenemos que hacer es normalizar los valores de las variables X_j de manera que $0 \leq X_j \leq 1$ con $j=1, \dots, d$. Además, en el modelo propuesto entenderemos que el cliente tiene mejor calificación U cuando más se aproxime a 0 y tendrá peor calificación cuando U se acerque más a 1.

Nótese que si $a_{1j}^* < 0$, entonces el sumando correspondiente es

$$-a_{1j}^* X_j$$

lo que hace que cuanto más posea el cliente de esta característica ($X_j \cong 1$) que influye negativamente en la concesión de un crédito, más penalizaremos la nota. Recíprocamente, cuanto menos posea de esta característica ($X_j \cong 0$) menor será la nota obtenida con respecto a esta misma.

Si $a_{1j}^* > 0$, entonces el sumando correspondiente será

$$a_{1j}^* (1 - X_j).$$

En esta situación, cuanto más posea el cliente de la característica ($X_j \cong 1$), la calificación que esta aporta a U es cercana a cero.

Por lo tanto, siguiendo este criterio de valoración, podemos deducir que $0 \leq U \leq 1$.

Uno de los objetivos planteados por Basilea II es la implantación de segmentación de créditos según el riesgo de concesión de los mismos. En nuestro modelo asociamos la medición del riesgo con la calificación que obtenga el cliente, en este caso, entendemos que cuanto menor sea la nota obtenida menor es el riesgo del capital prestado y viceversa. Esto nos conduce pues, a hacer una segmentación de las calificaciones obtenidas por los clientes mediante la aplicación del Algoritmo de Kohonen.

El proceso de aplicación del algoritmo está más detallado en el apartado siguiente y como se puede ver, al realizar la clasificación de las notas U podemos obtener un vector:

$$\mathbf{q}^* = [q_1^*, q_2^*, \dots, q_m^*]'$$

donde cada q_i^* representa la calificación hacia la que convergen las notas de los clientes próximos, y por lo tanto, decimos que esos están representados por la calificación q_i^* . Cada q_i^* representará a un segmento de riesgo quedando los extremos de los segmentos \tilde{q}_i configurados de la siguiente manera:

$$\tilde{q}_1 = 0, \quad \tilde{q}_k = (q_k^* + q_{k+1}^*)/2 \quad \text{para} \quad 2 \leq k \leq m, \quad \tilde{q}_{m+1} = 1$$

Al mismo tiempo cada q_i^* será el valor del riesgo de concesión de riesgo en el segmento al que representa y como $0 \leq U \leq 1$ diremos que cada valor q_i^* es la probabilidad de impago de los clientes que pertenecen a ese segmento de riesgo.

Podemos asegurar que el vector \mathbf{q}^* es único siempre que la función de densidad de las calificaciones obtenidas f_Z cumpla la propiedad de log-concavidad y esto se cumplirá cuando:

$$f_Z = f_{-a_{11}^* X_1} \dots f_{-a_{1s}^* X_s} \dots f_{a_{1s+1}^* (1 - X_{s+1})} \dots f_{a_{1d}^* (1 - X_d)}$$

Esto es, cuando la función de densidad de las calificaciones coincida con la convolución de las funciones de densidad de las calificaciones parciales.

2.3 Aplicación del Algoritmo de Kohonen en el modelo de Credit Scoring

Recordamos que estamos considerando un conjunto de aplicaciones de préstamos o prestatarios ya existentes en una institución bancaria a partir de los cuales se quiere crear una clasificación que le permita a la entidad disponer de un credit scoring y por lo tanto poder tomar decisiones de cobertura de capital. Así, ante la disyuntiva de conceder o no un préstamo a un nuevo cliente, el objetivo de la elaboración de nuestro credit scoring es obtener una calificación de este, de manera que podamos asociar esta calificación con el posible pago o impago del crédito solicitado y decidir si se le concede o no y en caso afirmativo medir el grado de riesgo que asume la entidad financiera en esta operación.

Si partimos de datos históricos sobre préstamos anteriores tendremos una sucesión de n calificaciones asociadas a los n clientes. Aplicaremos el Algoritmo de Kohonen a esta sucesión de calificaciones para poder realizar una partición de ellas en segmentos que nos reflejarán una tendencia de comportamiento similar de los clientes que pertenezcan a un mismo segmento.

Consideraremos que nuestra muestra aleatoria se extrae de una sucesión

$$\{U^{(n)}\}_{n \in N}$$

donde U es una variable aleatoria independiente con distribución continua μ . Nuestro objetivo es buscar una segmentación en función de su distribución de probabilidad. Para obtener esta segmentación utilizaremos el Algoritmo de Kohonen unidimensional. Este algoritmo está basado en un proceso de aprendizaje adaptativo y no supervisado.

El objetivo de este algoritmo es la obtención de segmentos de calificaciones para la sucesión $\{U^{(n)}\}_{n \in N}$ de manera, que cada uno de ellos represente un valor de probabilidad de impago, y así, cuando la calificación de un cliente pertenezca a un segmento determinado, podremos decir que su probabilidad de impago es la que está asociada a este segmento. Así, el objetivo es encontrar un vector de valores de calificación

$$\mathbf{q}^* = [q_1^*, q_2^*, \dots, q_m^*]'$$

de manera que la sucesión $\{U^{(n)}\}_{n \in N}$ queda dividida en m clases o segmentos de calificación y cada q_i^* es la calificación representativa de cada segmento y que debe estar asociada a la probabilidad de impago que el mismo representa.

Ahora, introducimos el Algoritmo de Kohonen, para explicar como obtenemos los vectores \mathbf{q}^* . Consideremos

$$\mathbf{q}^0 = [q_1^0, q_2^0, \dots, q_m^0]'$$

un valor inicial que representa una segmentación de las calificaciones, ahora tomaríamos la calificación de un cliente U_1 y mediante la aplicación de la fase competitiva del algoritmo,

$$\arg \min_{k=1 \dots m} |U^{(n+1)} - q_k^n|$$

mediríamos la distancia de esta con cada componente de \mathbf{q}^0 , y asociaríamos al cliente 1 con el q_i^0 cuya distancia haya resultado menor, pero ahora la componente q_i^0 se vería afectada en su valor mediante la aplicación de la fase cooperativa del Algoritmo de Kohonen.

$$q_k^{n+1} = q_k^n - \epsilon_{n+1} (q_k^n - U^{(n+1)})$$

donde $(\varepsilon_n)_{n \geq 1}$ es una secuencia de números reales comprendidos en el intervalo $[0,1]$. Usualmente se conoce a esta sucesión como ratio de aprendizaje.

Esta operación se realizaría con todos los valores de la sucesión $\{U^{(n)}\}_{n \in N}$ y obtendríamos \mathbf{q}^1 , así, diríamos que se ha realizado un epoch.

Si este proceso lo realizamos n veces (n epochs), obtenemos:

$$\mathbf{q}^n = [q_1^n, q_2^n, \dots, q_m^n]'$$

que representará la segmentación en la iteración $n \in N$.

Habremos obtenido el vector \mathbf{q}^* cuando $q^n = q^{n+1}$ y por lo tanto se cumple que $q^n = q^{n+1} = \mathbf{q}^*$

Los valores q_i^* son valores de calificación hacia los que convergen las calificaciones cercanas a ellas, por lo tanto, serán los representantes del valor de la calificación de cada segmento, quedando los extremos de los segmentos \tilde{q}_i configurados de la siguiente manera:

$$\tilde{q}_1 = 0, \quad \tilde{q}_k = (q_k^* + q_{k+1}^*)/2 \quad \text{para } 2 \leq k \leq m, \quad \tilde{q}_{m+1} = 1,$$

Cada q_k^* con $k=1, \dots, m$ sería el valor representativo de todas las calificaciones englobadas en el intervalo $\tilde{q}_k < q_k^* \leq \tilde{q}_{k+1}$.

2.4 Cálculo de las componentes de riesgo

El Nuevo Acuerdo de Basilea II obliga a las entidades bancarias que quieran utilizar un método interno de medición de riesgo a que este permita calcular unas determinadas componentes de riesgo, que son: la Probabilidad de Impago y la Pérdida dada de Impago, o bien la Pérdida Esperada.

Si se trata de un método pensado para exposiciones al detalle (pequeños préstamos), el valor de estas componentes de riesgo se puede calcular por segmentos de riesgo.

Una vez definido nuestro modelo pasamos a explicar a continuación como podemos hallar los valores de dichas componentes.

Probabilidad de impago

La probabilidad de impago o tasa de morosidad se define como la probabilidad de que el cliente entre en mora en un determinado periodo de tiempo.

Como q_i^* es el valor de convergencia de las calificaciones U comprendidas en el intervalo $[\tilde{q}_{i-1}, \tilde{q}_i]$, podemos decir entonces, que el conjunto de clientes agrupados en cada segmento $[\tilde{q}_{i-1}, \tilde{q}_i]$ tiende a comportarse como si hubieran alcanzado cada uno de ellos la calificación q_i^* .

Supongamos que $q_1^* \cong 0$, podríamos decir que los clientes del segmento $[\tilde{q}_0, \tilde{q}_1]$ son solventes con una probabilidad de impago igual a cero. En el caso contrario, con $q_m^* \cong 1$, los clientes comprendidos en el segmento $[\tilde{q}_{m-1}, \tilde{q}_m]$ tendrían una probabilidad de impago igual a 1.

Si esto es válido para los extremos, nos hace pensar que también lo es para los segmentos intermedios, puesto que el Algoritmo de Kohonen nos ratifica que el valor $q_i^* \in [\tilde{q}_{i-1}, \tilde{q}_i]$ es el valor

de convergencia del segmento. En nuestro caso, q_i^* representa calificaciones semejantes obtenidas por clientes de características parecidas y por lo tanto con una probabilidad de comportamiento semejante, comportamiento que el Algoritmo nos permite calcular y concretar en los valores q_i^* y por lo tanto definimos la probabilidad de impago para cada segmento como

$$P(D|U \in [\tilde{q}_{i-1}, \tilde{q}_i]) = q_i^*$$

Tanto la **Pérdida dada de impago (LGD)** como la **Pérdida Esperada (EL)** se desarrollan en el siguiente apartado del capítulo.

3. Aplicación y resultados

Se han realizado varios intentos para conseguir datos reales proporcionados por entidades bancarias para poder aplicar el modelo propuesto. El resultado de nuestra demanda ha sido totalmente infructuoso, puesto que se acogen al secreto de datos para la no concesión de los mismos.

Así, nuestra única posibilidad de hacer una aplicación práctica la encontramos en los datos del CD que incorpora el libro "Credit Scoring and its Applications" [15]. Para ello, ha sido necesario emplear los programas Excel, que ha sido utilizado en el proceso de manipulación de datos, con SPSS se han podido realizar los cálculos del análisis discriminante, y Matlab nos ha permitido aplicar el Algoritmo de Kohonen haciendo uso de la toolbox Neural Networks en el apartado One-dimensional self-organizing map.

Disponemos en principio de doce variables para explicar el pago o impago de los créditos concedidos por una entidad bancaria a 897 clientes ($n=897$) de los que conocemos que 632 préstamos han sido amortizados, mientras que los 265 restantes han resultado insolventes. Para trabajar más cómodamente, asociaremos a cada variable con una abreviatura que utilizaremos a partir de ahora para referirnos a ellas, así tendremos:

Tabla 1: Abreviatura de las variables utilizadas originalmente

Variable	Abreviatura
Año de nacimiento	Año
Nº de niños	Niñ.
Personas dependientes	Pdep.
Tenencia de teléfono	Telf.
Renta del cónyuge	Rtacy.
Renta del cliente	Rtacl.
Valor del hogar	Vhg.
Hipoteca pendiente	HPpen.
Gastos en hipoteca o alquiler	Gtohaq
Gastos en préstamos	Gtopr.
Gastos en compras a plazos	Gtocpz
Gastos en tarjetas de crédito	Gtoctr.

Fte: Elaboración propia

Estos datos los sometemos al análisis discriminante en el programa SPSS y obtenemos los siguientes resultados:

Tabla 2: Anovas Univariantes

Variable	F	Sig=p
Año	5.077	0.024
Niñ.	0.578	0.447
Pdep.	0.613	0.434
Telf.	0.828	0.363
Rtacy.	1.327	0.250
Rtacl.	8.169	0.004
Vhg.	0.022	0.881
HPpen.	0.404	0.525
Gtohaq	0.931	0.335
Gtopr	0.054	0.816
Gtospz	3.955	0.047
Gtotcr.	1.327	0.250

Fte: Elaboración propia

Tabla 3: Valores de los estadísticos de significación de las variables sometidas a estudio.

Variable	Lambda de Wilks	F	Sig=p
Año	0.987	16.101	0.000
Niñ.	0.998	1.875	0.171
Pdep.	1	0.224	0.636
Telf.	0.998	2.291	0.130
Rtacy.	0.996	4.606	0.032
Rtacl.	0.966	43.614	0.000
Vhg	0.998	2.571	0.109
HPpen.	0.998	2.18	0.140
Gtohaq	0.995	5.633	0.018
Gtopr	0.999	1.278	0.258
Gtospz	0.995	6.623	0.010
Gtotcr	0.996	4.606	0.032

Contraste Glogal	0.977	Significatividad en el modelo $\chi^2 = 30.795$	0.02
------------------	-------	----------------------------------------------------	------

Fte: Elaboración propia

Antes de sacar conclusiones de los resultados obtenidos recordamos las definiciones de los estadísticos que hemos utilizado:

Lambda de Wilks: Contraste de significación multivariado, algunas veces denominado el estadístico de U. Lambda toma valores entre 0 y 1. Los valores próximos a 0 indican que las medias de los grupos son diferentes, mientras que los valores próximos a 1 indican que las medias de los grupos no son diferentes, que sea igual a 1 indica que todas las medias son iguales.

F: Cociente entre dos medias cuadráticas. El estadístico F solo puede tomar valores positivos o cero. Así, cuando la hipótesis nula es cierta, esperamos que F tome valores próximos a uno. A medida que

las medias están más separadas entre sí, el valor F se hace mayor. Los valores grandes de F constituyen una prueba fehaciente en contra de la hipótesis nula, inclinándonos a pensar que la hipótesis correcta es la alternativa.

sig=p: El p-valor o nivel de significación empírico del contraste es el dato obtenido a partir del valor del estadístico de contraste, en las observaciones que corresponden a la realización de la muestra de tamaño n extraída de una población, y que nos informa sobre cuál sería el nivel de significación más pequeño que nos hubiera permitido rechazar la hipótesis nula. Se rechazará la hipótesis nula si el p-valor es menor o igual al nivel de significación adoptado por el experimentador. De manera habitual, se considera significativo un valor menor que 0.05.

Observando los estadísticos la primera conclusión que obtenemos por los valores de la Lambda de Wilks es que en nuestros datos (n=897), no hay mucha diferencia en las características de los clientes que han pagado de los que no han pagado, lo cual supone una dificultad mayor. Con la ayuda de los otros dos estadísticos tomamos como características significativas aquellas con valor de Sig=p<0.05 que además coinciden con un mayor valor del estadístico F, y dentro de un margen con un menor valor de la Lambda de Wilks.

Para ratificarnos en la toma de decisión de las variables significativas sometemos a las mismas al análisis denominado Anova de un factor que genera un análisis de varianza de un factor para una variable cuantitativa respecto a una única variable de factor (la variable independiente). El análisis de varianza se utiliza para contrastar la hipótesis de que varias medias son iguales. Los resultados que obtenemos quedan reflejadas en la Tabla 2.

La hipótesis nula que se plantea al aplicar Anova es que las medias son iguales, el programa SPSS responde con un estadístico F relativamente grande y con p pequeño para rechazar la hipótesis nula. Fijándonos entonces en la tabla 2 solo podríamos trabajar con tres variables, Año, Rtacl, y Gtopz. Estos resultados nos vienen a confirmar que los datos disponibles no son buenos, que hay mucha igualdad en los valores de las variables que caracterizan a los clientes solventes de los que no lo son, que dificultan la aplicación del análisis discriminante, o bien, que los datos no se ajustan a un modelo lineal. Sin embargo, la no bondad de estos datos nos puede ayudar a ratificar la justificación de aplicar un segundo método que nos permita una mejor clasificación de los clientes.

Trabajar con solo tres variables reduciría mucho la posibilidad de separación de clientes, por lo tanto a pesar de conocer que el resto de variables no aporta demasiado a la solución del problema, y aún sabiendo que no son significativas vamos a añadir las tres variables siguientes en significación discriminante y que en principio la Lambda de Wilks nos daba como aceptables¹.

Trabajaremos definitivamente con las siguientes variables significativas obtenidas en la tabla 2 cuya matriz de correlación es:

Tabla 4: Matriz de covarianzas de las variables significativas

	Rtacl.	Año	Rtacy.	Gtotcr.	Gtopz	Gtohaq
Rtacl.	1	-0.085	-0.002	-0.002	0.142	0.383
Año	-0.085	1	-0.118	-0.118	-0.014	0.046
Rtacy.	-0.002	-0.118	1	1	0.021	0.131
Gtotcr	-0.002	-0.118	1	1	0.021	0.131
Gtopz	0.142	-0.014	0.021	0.021	1	0.017
Gtohaq	0.383	0.046	0.131	0.131	0.017	1

Fte: Elaboración propia

Si nos fijamos en los resultados obtenidos lo primero que llama nuestra atención es la similitud en el valor de los estadísticos de las variables Renta del cónyuge y Gastos en tarjetas de crédito, comprobando además que existe una correlación 1 entre ellas, esto significa que ambas tienen la misma capacidad explicativa en el pago o impago de los préstamos y por lo tanto descartamos una de

¹ La hipótesis nula que se propone es que la Lambda de Wilks es igual a 1

ellas, y hemos optado por no utilizar la variable Gastos en tarjetas de crédito, y nos quedamos definitivamente con cinco variables para que nos expliquen los pagos o impagos de estos créditos. Hay que hacer notar también que las correlaciones entre las demás variables no son muy altas lo que nos da idea de la fuerte componente ortogonal de las características o variables explicativas.

El siguiente paso consistirá en determinar cual es el grado de significatividad de estas cinco variables, esto es, cuál es el grado de incidencia de cada una de ellas en la explicación de la variable dependiente. Si nos fijamos en las tablas anteriores y en los valores de los estadísticos, serán más significativas aquellas variables que tengan una Lambda de Wilks menor, un estadístico F mayor y un estadístico p menor. Pero con estos datos solo obtendríamos una clasificación ordinal y lo que buscamos es una valoración cardinal de su faceta explicativa, para ello contamos con los valores de tres coeficientes diferentes:

1. Coeficientes estandarizados de las funciones canónicas discriminantes
2. Coeficientes de las funciones canónicas discriminantes
3. Coeficientes de las funciones lineales de Fisher.

Para ello elegimos los valores de los coeficientes enumerados en primer lugar como ya queda explicado anteriormente.

Al aplicar sobre los datos de nuestra muestra el programa SPSS, además de los datos estudiados anteriormente, también obtenemos el valor de los coeficientes estandarizados de las funciones canónicas discriminantes.

Así, los valores, ya ordenados de mayor a menor, que se obtienen son los siguientes:

Tabla 5: Valoración del grado de incidencia de las variables en el pago de los préstamos

Variable	a_{1j}^*	Valor
Rtacl.	a_{16}^*	0.83
Año	a_{11}^*	0.58
Rtacy	a_{15}^*	0.33
Gtocz.	$a_{1,11}^*$	0.182
Gtohaq	a_{19}^*	-0.116

Fte: Elaboración propia

Nos llama la atención que la variable Gastos en compras a plazos tenga una incidencia positiva en el pago de los préstamos, ya que cabría esperar que cuanto mayor fuera el endeudamiento en el que incurre el cliente, menor fuera su capacidad de respuesta al pago del préstamo.

Como no tenemos ninguna información adicional en la que sustentar una explicación lo tomamos tal y como nos viene dado, aunque nuestra interpretación ante los resultados sería que esta variable indica la respuesta efectiva de los clientes ante el pago a plazo.

Además de toda esta información, el análisis discriminante también aporta una clasificación previa de los clientes que recogemos en la siguiente tabla, siguiendo la proporción original:

Tabla 6: Pronóstico de préstamos pagados e impagados utilizando Análisis Discriminante con proporciones iniciales del 70.5% y 29.5% respectivamente.

	Pronosticados buenos	Pronosticados malos	Total
Porcentaje buenos	97%	3%	100%
Porcentaje malos	88.9%	11.1%	100%

Fte: Elaboración propia

Estos resultados explican que del total de la muestra están clasificados correctamente el 74,4% de los casos.

Si la probabilidad a priori de impago es del 50% obtendríamos la siguiente clasificación.

Tabla 7: Pronóstico de préstamos pagados e impagados utilizando Análisis Discriminante

	Pronosticados buenos	Pronosticados malos	Total
Porcentaje buenos	41.5%	58.5%	100%
Porcentaje malos	26%	74%	100%

Fte: Elaboración propia

En este caso del total de la muestra quedarían clasificados correctamente solo el 48.8% de los casos, aunque hay que advertir que el porcentaje de acierto en los clientes insolventes es mucho mayor ya que aumenta del 11.1% al 74%

Con los datos iniciales se testaron varios modelos estadísticos como logit y probit y el que mejores resultados aportaba fue el análisis discriminante².

Aún así, vemos que con nuestra muestra es fácil detectar los clientes solventes, pero no así, los insolventes y esto se debe, como ya hemos indicado anteriormente, a la similitud de las características entre buenos y malos clientes.

Nuestro objetivo radica en intentar conseguir mejorar estos resultados con la aplicación del modelo que proponemos.

Con toda la información que tenemos razonamos de la siguiente manera para calificar a los clientes:

1. Cuanto mayor sea el valor de a_{ij}^* de una característica, más relevante será para el pago del préstamo y su efecto sobre el mismo estará relacionado directamente con el signo que le acompañe.
2. Cuanto más posea el cliente de aquella característica que afecta positivamente al pago, mejor puntuación debe obtener el cliente y viceversa.
3. Queremos calificar al cliente en un intervalo $[0,1]$ donde los mejores clientes tenderán a una nota de 0 y los no pagadores se acercarán al 1.

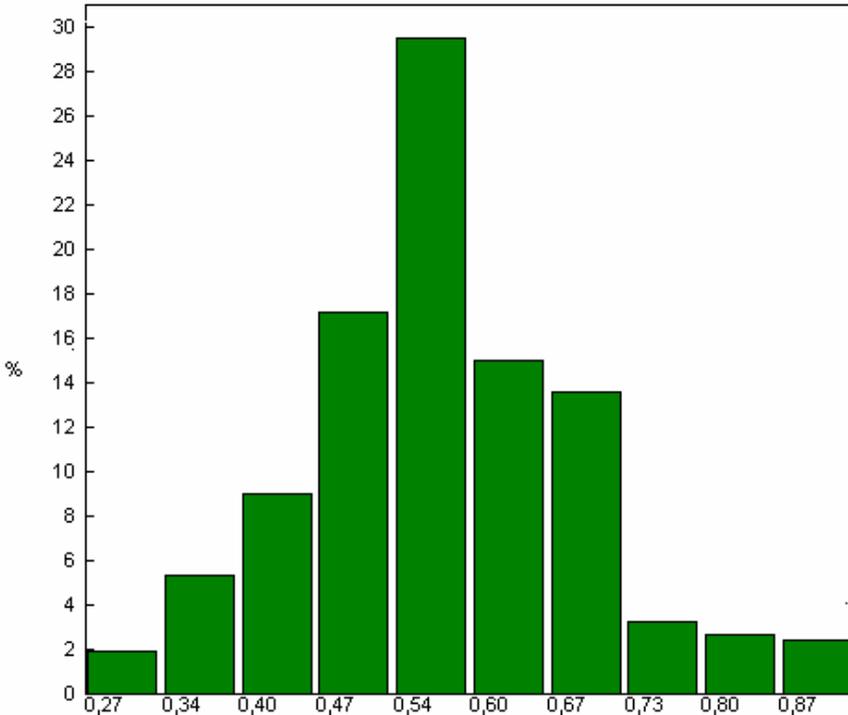
Con esto llegamos a la conclusión de que la calificación de cada cliente U vendrá dada por la expresión definida con anterioridad.

A continuación tomamos la sucesión de datos U que hemos obtenido como calificaciones de los clientes y le aplicamos el Algoritmo de Kohonen para obtener los segmentos de calificación que nos indiquen un mismo valor de riesgo, y con ello, calcular los componentes de riesgo definidos por Basilea II. Para poder hacer esto deben seguirse los siguientes pasos:

Gráfica de la distribución de las calificaciones de los clientes U

² Si el lector desea aplicar este modelo en otra muestra, se recomienda testar los datos con diferentes métodos estadísticos y elegir el que aporte mejores resultados

NOTA FINAL : Nodo 0



- 1) La función de distribución de U debe cumplir la propiedad de log-concavidad, esto es, si calculamos el logaritmo de la función de distribución esta debe ser cóncava. Como no conocemos la función de distribución de esta variable hacemos una exploración por SPSS y como se ve en la siguiente gráfica podríamos asegurar que la sucesión de calificaciones se distribuye con una función de distribución que cumple la log-concavidad.
- 2) Tenemos que decidir el número de segmentos o clases, que denominaremos m , en los que van a quedar divididas las notas U obtenidas por los clientes, y en nuestro caso decidimos, mirando la distribución, probar con $m=4$. La razón de esta elección y no otra es, que como los segmentos nos van a clasificar a los clientes por riesgo de impago, el orden de las clases nos indicarían clientes de riesgo nulo, clientes de bajo riesgo, clientes de riesgo medio y clientes de alto riesgo.
- 3) El siguiente paso entonces, es aplicar el Algoritmo de Kohonen, utilizando Matlab, a la distribución de calificaciones $U \in [0,1]$, indicándole que $m=4$, y el programa asigna de forma aleatoria unos valores iniciales de manera que el segmento $[0,1]$ queda dividido en cuatro clases. Así, el programa comienza a ejecutar el Algoritmo tal y como se explica en el capítulo 4 de este trabajo y obtenemos los siguientes resultados $\mathbf{q}^* = [0.4772, 0.5368, 0.6253, 0.7006]$. Por lo tanto los segmentos de agrupación de calificaciones serían: $[0, 0.507]$ $]0.507, 0.58105]$ $]0.58105, 0.66295]$ y $]0.66295, 1]$
- 4) Una vez obtenida la solución del problema, tenemos que comprobar si esta es correcta o no lo es. La solución no sería correcta si las clases o segmentos obtenidos no fueran representativos de los datos que contienen, es decir, podríamos encontrarnos con dos casos en los que la solución no fuera acertada.
 - a) Que el número de clases elegido sea menor que el que se debería haber cogido, y esto implicaría que habría datos que podrían dar más información, y en nuestro caso, representarían por ellos mismos otro segmento de riesgo y los estamos obligando a pertenecer a un nivel equivocado, y

b) Que el número de segmentos elegido es mayor que el que se debería haber tomado, y por lo tanto, estamos segmentando el riesgo en más niveles de los que verdaderamente se están dando en la entidad bancaria.

La solución a este problema lo encontramos en el mismo programa Matlab mediante la aplicación de la Función de Aprendizaje de pesos³.

Si los pesos han aprendido bien, es decir, si la solución es correcta, la Función de Aprendizaje nos proporciona una matriz ⁴ 4×4 en cuya diagonal principal tendremos unos y en el resto ceros, y ya adelantamos que este es el resultado obtenido, por lo tanto, queda certificada la bondad de la solución conseguida. Si el número de clases no hubiese sido el correcto aparecería un cero en la diagonal principal, y el uno que correspondería en ese lugar se trasladaría a aquella fila donde hubiera sido más significativo agrupar esa clase, esto nos estaría indicando que m debe ser menor que cuatro.

Por otra parte, repetimos el ejercicio suponiendo m=5 y los resultados obtenidos fueron:

$$q^* = [0.3117 \ 0.482, 0.5418, 0.628, 0.7093]$$

pero al aplicar la Función de Aprendizaje la matriz 5×5 obtenida fue:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

lo que nos indicaba que los datos del primer segmento no eran significativos y que su tendencia era la misma que la del segundo segmento como así queda comprobado en la solución del problema.

A continuación presentamos una tabla donde representamos los pagos e impagos en cada uno de los segmentos

Tabla 8: Frecuencia absoluta de pagos e impagos en cada intervalo de calificación de los clientes

Segmento	Pagos	Impagos
[0, 0.507]	219	2
]0.507, 0.58105]	220	50
]0.58105, 0.66295]	127	78
]0.66295, 1]	66	135

Fte: Elaboración propia

Hemos representado las calificaciones obtenidas por los clientes en un gráfico utilizando el programa Excel que aparece en el anexo.

En este gráfico encontramos en el eje Definición solo dos valores, 0 y 1. El 0 representa aquellos clientes que han pagado su préstamo y el 1 los clientes que no pagaron su préstamo.

³ Los pesos son los valores obtenidos como solución

⁴Hemos decidido que m=4

En el eje de Calificación representamos la puntuación obtenida por los clientes según el modelo desarrollado, que como hemos indicado está definida en el intervalo [0,1]. Si nos detenemos en el gráfico vemos que hay algunos valores relevantes en la calificación que vamos a ir detallando paulatinamente.

En el segmento] 0.66295, 1] hay un valor importante que es 0.71443 puesto que no se produce ningún pago en clientes con $U > 0.71443$ que nos permite establecer una clasificación de los clientes que presentamos en la siguiente tabla.

Tabla 9: Pronóstico de préstamos pagados e impagados utilizando el Algoritmo de Kohonen tomando como referencia el último segmento

	Pronosticados buenos	Pronosticados malos	Total
Porcentaje buenos	100% si $U < 0.71443$	0% si $U < 0.71443$	100%
Porcentaje malos	64% si $U < 0.7443$	36% si $U > 0.71443$	100%

Fte: Elaboración propia

Si estudiamos los resultados obtenidos en el otro segmento extremo [0, 0.507] vemos que todos los clientes que han obtenido su calificación en este segmento han resultado solventes con excepción de dos cuyas calificaciones son mayores que 0.4916, esto es, muy cercanas al valor de cierre del segmento 0.507.

El valor aportado por el algoritmo a este segmento es 0.4772, calificación que se puede interpretar como valor indicativo a partir del cual los clientes pueden comenzar a ser insolventes y al mismo tiempo la entidad financiera siempre podrá esperar de todos aquellos clientes que obtengan una calificación inferior que sean solventes. Si tomamos como valor de referencia el del cierre del segmento 0.507, podemos elaborar otra tabla resumen de los porcentajes pronosticados que serían los siguientes:

Tabla 10: Pronóstico de préstamos pagados e impagados utilizando el Algoritmo de Kohonen tomando como referencia el primer segmento

	Pronosticados buenos	Pronosticados malos	Total
Porcentaje buenos	34.7% si $U < 0.507$	65.3% si $U > 0.507$	100%
Porcentaje malos	0.7% si $U < 0.507$	99.3% si $U > 0.507$	100%

Fte: Elaboración propia

Si recordamos los datos obtenidos solo con la aplicación del análisis discriminante podemos establecer una comparación de los resultados de la Tabla 6 con los obtenidos en la Tabla 9 ya que las probabilidades a priori (70%) son muy cercanas a la nota de corte $U = 0.71443$. Así, vemos que se ha obtenido una mejora en la predicción de malos clientes frente a la obtenida por el análisis discriminante que nos pronosticaba el 11,1%.

Además en este mismo tramo del gráfico, no encontramos ningún cliente que haya pagado su préstamo y que sea calificado como mal cliente, es decir nuestro error es del 0% mientras con el análisis discriminante se cometía un error del 3%. Lo que implica además un porcentaje de acierto global del 81% frente al 74,4% que obteníamos solo con el análisis discriminante.

Si seguimos analizando resultados en este último segmento apreciamos que este valor de corte entre pagos e impagos (0.71443) es muy próximo al valor aportado por el algoritmo de Kohonen 0.7006, que se puede considerar como un valor de alarma para la entidad financiera a la hora de conceder préstamos a clientes cuya calificación oscile entorno a ese valor.

De la misma manera establecemos una comparación de los datos obtenidos en la Tabla 7 con los de la Tabla 10 ya que la nota de corte $U=0.507$ es coincidente con la probabilidad a priori (50%) y vemos que se ha producido una gran mejora en el pronóstico de clientes insolventes que pasa de un error del 26% al 0.7%.

Podemos entonces resumir la información que se puede obtener del gráfico en la tabla siguiente:

Tabla 11: Frecuencias relativas de pagos e impagos por cada intervalo de calificación del cliente

Segmento	Pagos	Impagos
[0, 0.507]	34.7%	0.7%
]0.507, 0.58105]	34.8%	18.9%
]0.58105, 0.66295]	20%	29.4%
]0.66295, 1]	10.5%	51%

Fte: Elaboración propia

Estos resultados nos sugieren que la población de esta entidad bancaria es bastante homogénea ya que el 53% está valorada en los intervalos intermedios,]0.507, 0.58105] y]0.58105, 0.66295] y tomando valores globales de los clientes situados en ellos, el 73% ha resultado solvente y el 27% incurre en impago. Por lo tanto sería en este tramo de calificación donde el banco tendría que afinar su toma de decisiones ya que hay que resaltar que las calificaciones intermedias de estos segmentos aportadas por el Algoritmo de Kohonen son 0.5368 y 0.6253, es decir, calificaciones cercanas a la media pero con tendencia a la insolvencia puesto que la superan.

Nuestro último objetivo es este trabajo es conseguir una relación entre las calificaciones obtenidas por los clientes y los componentes de valoración del riesgo que son la Probabilidad de impago, la Pérdida dada de impago y la Pérdida Esperada.

Para ello tomamos los valores obtenidos por el Algoritmo de Kohonen y que configuran el vector \mathbf{q}^*

$$\mathbf{q}^* = [0.4772, 0.5368, 0.6253, 0.7006]$$

Cada uno de estos valores reflejan las concentraciones de valor de las calificaciones en cada uno de los segmentos y es por esto que en nuestro modelo vamos a asociar la PD (probabilidad de impago), que se define como la probabilidad de que el cliente entre en mora en un determinado periodo de tiempo, con los valores q_i^* .

Por lo tanto

$$PD = [0.4772, 0.5368, 0.6253, 0.7006]$$

En cuanto al cálculo de la pérdida dada de impago que representa el ratio entre las pérdidas efectivas incurridas, como consecuencia de la entrada en mora, y la Exposición, tenemos que recordar que consideramos los préstamos concedidos en un mismo producto crediticio, y que estos son homogéneos en cantidad, esto es, a todos los clientes se les ha concedido un préstamo de la misma cantidad que

suponemos igual a 1, por lo tanto $C = \sum_{l=1}^m C_l = 897$ es el capital total expuesto, por segmentos la cantidad prestada ha sido la siguiente:

Tabla 12: Capital concedido en préstamos en cada segmento

Segmento	C_l
[0, 0.507]	221
]0.507, 0.58105]	270
]0.58105, 0.66295]	205
]0.66295, 1]	201

Fte: Elaboración propia

Otro de los componentes de riesgo es la Pérdida dada de impago (LGD) también denominada Severidad que representa el ratio entre las pérdidas efectivas incurridas, como consecuencia de la entrada en mora, y la Exposición. Sabiendo esto, a partir de los datos de la tabla 9, podemos decir que la Pérdida dada de Impago (LGD) en cada uno de los segmentos de riesgo será la siguiente:

Tabla 13: Valor de la Severidad en cada intervalo de riesgo

Segmento	LGD
[0, 0.507]	0.9%
]0.507, 0.58105]	18.5%
]0.58105, 0.66295]	38%
]0.66295, 1]	67%

Fte: Elaboración propia

En cuanto al último componente de riesgo, el cálculo de la Pérdida Esperada (EL) cuyo concepto coincide con la media matemática de la pérdida crediticia, viene determinada por Basilea II⁵, y se define como:

$$EL=PD \times EAD \times LGD$$

En nuestro caso el resultado que obtenemos es que por cada 100 unidades de préstamos concedidos en cada segmento la Pérdida Esperada es:

Tabla 14: Valor de la Pérdida Esperada por cada 100 unidades monetarias prestadas

Segmento	EL
[0, 0.507]	0.95 u.m.
]0.507, 0.58105]	26 u.m.
]0.58105, 0.66295]	48 u.m.
]0.66295, 1]	94 u.m.

Fte: Elaboración propia

⁵ Basilea II propone esta fórmula de la Pérdida Esperada para exposiciones empresariales. Al no encontrar ninguna propuesta alternativa para exposiciones al detalle, proponemos la posibilidad de utilizar esta fórmula para el cálculo de la Pérdida Esperada en cada segmento de riesgo.

4.CONCLUSIONES

A pesar de no contar con los datos más idóneos para la resolución del problema planteado se ha podido llegar a las siguientes conclusiones:

1. Como apuntábamos en el apartado 4 uno de nuestros objetivos era comprobar si los resultados obtenidos en un Credit Scorig con la conjunción de modelos estadísticos tradicionales y redes neuronales [1] eran mejores que los obtenidos con la aplicación de estos modelos por separado y hemos visto en el desarrollo del ejercicio que así es. El análisis discriminante permite seleccionar las variables relevantes, propiedad que no poseen las redes neuronales, y la incorporación del Algoritmo de Kohonen permite conseguir una mejor diferenciación de los clientes solventes e insolventes que la obtenida solo con la aplicación del análisis discriminante.
2. Los resultados obtenidos a partir del Algoritmo, \mathbf{q}^* , son los valores de las calificaciones de los clientes que identifican cada segmento de riesgo, y estos valores, dependen al mismo tiempo, de los coeficientes estandarizados obtenidos en el Análisis Discriminante, por lo tanto los resultados que obtenemos se derivan de un modelo mixto compuesto por redes neuronales y análisis discriminante.
3. La solución \mathbf{q}^* obtenida es única y estable. La unicidad y estabilidad vienen confirmadas por la convergencia del Algoritmo de Kohonen en el caso unidimensional [17], que es el que hemos aplicado en este trabajo.
4. Según el modelo desarrollado decidimos o no conceder un préstamo en función de la calificación q_i^* . Si estamos ante un prestatario cuyas características indican que es un buen cliente, q_i^* tenderá a cero y viceversa, por lo tanto podemos definir la probabilidad de impago como: $PD=q_i^*$. En el ejemplo utilizado, hemos obtenido que los valores inferior y superior del vector \mathbf{q}^* nos permiten descartar con una alta fiabilidad los clientes solventes e insolventes respectivamente. Cómo podía esperarse, son los valores intermedios del vector \mathbf{q}^* los que más incertidumbre generan a la entidad bancaria a la hora de decidirse sobre la concesión o negación del préstamo, ya que estamos hablando de una probabilidad de impago entre el 53% y el 62%. El hecho de poder segmentar el riesgo de concesión de créditos es una conclusión que nos parece puede ser aplicada como modelo para las exigencias del Nuevo Acuerdo de Basilea en exposiciones al detalle, ya que los valores de riesgo de impago se desglosan en intervalos o segmentos, como exige el Nuevo Acuerdo y hay que advertir que cuanto mayor sea el número de intervalos, es decir, cuanto mayor sea m la precisión en el cálculo del ratio de impago es mayor, siempre que la función de aprendizaje del algoritmo nos certifique que las m clases son significativas.
5. También concluimos, que podemos obtener otra de las variables o conceptos descritos por Basilea II como una componente de riesgo a calcular para cada segmento obtenido y esta es la pérdida dada de impago y esta es la pérdida dada de impago cuya forma general ya indicada en el capítulo sería: $LGD=(\overline{C}_i)/(C_i)\times 100$
6. Y por último, y teniendo en cuenta el supuesto de que podemos aplicar la obtención de la pérdida esperada, definida por Basilea II para exposiciones empresariales, a exposiciones al detalle en nuestro trabajo práctico los resultados han sido:

Segmento	EL
[0, 0.507]	0.95 u.m.
]0.507, 0.58105]	26 u.m.
]0.58105, 0.66295]	48 u.m.
]0.66295, 1]	94 u.m.

Donde observamos que la pérdida esperada va creciendo junto con la probabilidad de impago.

7. Por todo lo expuesto en las conclusiones anteriores, creemos haber elaborado un Credit Scoring que cumple las directrices de Basilea II para pequeños préstamos. Su incorporación en la banca lleva la necesidad de implementar nuevas técnicas, lo que supone incurrir en un coste técnico y de profesionales tanto en el ámbito de departamentos de riesgo de los bancos, como por parte del supervisor encargado de confirmar la idoneidad de estos métodos. Desde estas páginas, queremos hacer un llamamiento a la banca privada para una mayor colaboración Universidad-Banca. Este trabajo hubiera podido ser más concluyente en su parte práctica, si los bancos y cajas de ahorros a los que fueron solicitados datos anónimos, hubieran colaborado, y estas entidades podrían contar con un trabajo de aplicación a sus préstamos con las características exigidas por Basilea II.

BIBLIOGRAFÍA

- [1] Altman, Edward I., Giancarlo Marco y Franco Varetto.(1994). "Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (The Italian Experience)." *Journal of Banking and Finance* 18 , pp.505-29.
- [2] : Benain, M. Fort, J.C., y Pagès G., (1998). "Convergence of the one-dimensional Kohonen Algorithm". *Adv. Appl. Prob.* 30, 850-869.
- [3] : Capon, N. "Credit scoring systems: A critical analysis". *J. Marketing*, 46, 82-91. (1982)
- [4] : Churchill, G., Nevin, J., Watson, R. "The role of credit scoring in the loan decision" *Credit World*, March, 6-10. (1977)
- [5] : Durand, D. "Risk Elements in Consumer Instalment Financing", National Bureau of Economic Research. New York. (1941)
- [6] : Fisher, R. "The use of multiple measurements in taxonomic problems", *Ann. Eugenics.* 7, 179-188. (1936).
- [7] : Freed, N., Glover, F. "A linear programming approach to the discriminant problem". *Decision Sci.*, 12, 68-74. (1981 A)
- [8] : Galindo J., Tamayo P., (1997). "Credit Risk Assessment using Statistical and Machine Learning Methods as an Ingredient for Financial Intermediaries Risk Modeling". Departamento de economía de la Universidad de Harvard y Thinking Machines Corp.
- [9] : www.bis.org/publ/bcbsca.htm. Nuevo Acuerdo de Basilea II
- [10] : Lewis, E. "An Introduction to Credit Scoring". Athena Press, San Rafael, CA. (1992).
- [11] : Myers, J., Forgy, E. "The development of numerical credit evaluation systems". *J.Amer.Statist. Assoc.*, 58, 799-806. (1963).
- [12] : Mester, L.J., (1997). "What's the point of Credit Scoring?". *Business Review*, Septiembre/Octubre 1997, Federal Reserve Bank of Philadelphia
- [13] : Roszbach, K., (1998). "Bank lending policy, credit scoring and the survival of loans", Escuela de economía de Estocolmo,
- [14] : Smalley, O., Sturdivant, F. "The Credit Merchants: A History of Spiegel". Southern Illinois University Press, Carbondale, IL. (1973).
- [15] : Thomas, L., Edelman, D., Crook, J. "Credit Scoring and its Applications" SIAM monographs on mathematical modeling and computation. (2002).
- [16] : Weingartner, H. "Concepts and utilization of credit scoring techniques". *Banking*, 58, 51-53. (1966)
- [17] : Fort, J.C., y Pagès G., Convergence on stochastic algorithms: from the Kushner & Clark theorem to the Lyapunov functional method. *Adv. Appl. Prob.*, 28, 1072--1094.(1996).

Anexo de calificaciones

