

Comparing Mutual Information and Delta Test for Variable Selection to Determine Client Confidence in e-Bank

F. Liebana

Marketing and Market Research Department. University of Granada. Spain
{franlieb}@ugr.es

A. Guillen, L.J. Herrera, H. Pomares, I. Rojas

Computer Technology and Architecture Department. University of Granada. Spain
{aguillen}@atc.ugr.es

Abstract – The new technological achievements that have been developed in the last decade are making a change in the profile and behaviour of Internet users and electronic banking. The implementation of this service and the significant increase of its use make mandatory its use for the finance sector and to make investments in order to satisfy and keep the clients by improving their confidence and loyalty. Electronic Bank has become one of the elements that distinguish one bank from another that can make a difference in a certain type of client. The purpose of this paper is to contribute to the research in this important area by applying novel machine learning techniques and algorithms in order to identify client's profiles as well as asses to apply strategies based on the data that a bank has available. Concretely, starting from a data base, Mutual Information and Delta Test will be applied to identify the most significant variables that determine the global satisfaction with the bank as well as the use of neural networks to predict the satisfaction level.

Keywords – Variable Selection, Mutual Information, Delta Test, RBFNN, e-Bank.

Resumen – Los desarrollos tecnológicos que se han producido en la última década han provocado un cambio en el perfil y comportamiento de los usuarios de Internet y Banca Electrónica. La implementación de este servicio y el importante crecimiento que ha supuesto su uso ha obligado al sector financiero a realizar inversiones que satisfagan las expectativas de sus clientes, favoreciendo la lealtad de los mismos. La Banca Electrónica se ha convertido en un elemento diferenciador entre las entidades financieras de manera que éstas tratan de incrementar el nivel de satisfacción de sus clientes cumpliendo con las expectativas que éstos manifiestan. El propósito de este trabajo es contribuir al campo de la investigación de la satisfacción desde la perspectiva del usuario de Banca Electrónica, campo de gran interés para la estrategia de negocio de las entidades financiera empleando para ello técnicas y algoritmos del área del aprendizaje máquina que permite la identificación de perfiles de clientes así como estrategias de actuación sobre los mismos a partir de variables estratégicas de tipo socio-demográficas, económico- financieras y comportamentales que la propia entidad dispone en sus bases de datos históricas. Concretamente, se han utilizado la información mutua y el test Delta para poder identificar las variables más significativas y, posteriormente, se ha diseñado una red neuronal capaz de predecir con relativa precisión el grado de satisfacción global.

Palabras Clave – Selección de variables, Información Mutua, Test Delta, RBFNN, Banca electrónica.

1.. Introduction

La satisfacción del consumidor se ha convertido en un factor de especial relevancia en la etapa de evaluación de una experiencia de compra, consumo o uso de un producto o servicio, y en consecuencia será vital para conocer las respuestas que el consumidor genere en el largo plazo [1–5]. En el sector financiero la importancia de la satisfacción también se ha incrementado principalmente por el nivel de competencia existente y por los desarrollos tecnológicos que se han sucedido en los últimos años. Los productos y servicios bancarios son muy similares entre todas las entidades financieras, de ahí que la mayoría de éstas traten de justificar su diferenciación y preferencia por parte de los clientes a través del nivel de satisfacción respecto a la prestación de servicios que reciben; y es la Banca Electrónica uno de los que mayores ventajas reporta al cliente financiero. Tradicionalmente, la distribución de los productos y servicios financieros se ha realizado en torno a la oficina bancaria por su cercanía al cliente, el elevado número de servicios que se realizaban en ella, el valor añadido que el cliente recibía en la oficina y la importancia que tenía la oficina en la toma de decisiones sobre los clientes, pero a pesar de ello, este canal ha comenzado a ser sustituido por uno más ágil y dinámico tal y como se desprende de los datos del World Retail Bank Report de 2010 donde la Banca Electrónica ya supone el 17 % del total de distribución financiera (incremento de un 70 % en tan solo cinco años) frente al 67 % del canal oficina bancaria (reducción del 28 % en cinco años). Desde la década de los noventa hasta la actualidad la Banca Electrónica está convirtiéndose en uno de los principales canales de distribución para las entidades financieras [6, 7]. En la actualidad la mayoría de las entidades ofrecen a sus clientes la posibilidad de acceder a la mayoría de sus servicios por este canal. La Banca Electrónica, por tanto, se ha convertido en un servicio determinante para conseguir la fidelización de sus clientes a través de la satisfacción de los mismos consiguiendo una mayor vinculación, de ahí su importancia manifiesta para cumplir las expectativas que los usuarios manifiestan [8–10] y una importante fuente de ahorro en costes para la propia entidad [11, 12], de ahí que conocer

la satisfacción de los usuarios a través de este canal sea considerada por las entidades una información estratégica para la gestión del canal de venta on line. El estudio de la satisfacción tiene sus orígenes en las investigaciones de [13] y [14] donde se analiza la influencia de las expectativas sobre la satisfacción. Estos estudios son considerados el punto de partida del interés científico hacia la satisfacción, sus orígenes y consecuencias [?, 15], a partir de estos han sido múltiples las aportaciones a la literatura científica, si bien es cierto que uno de los modelos más analizados ha sido el Modelo de Desconfirmación de Expectativas, (Expectation Disconfirmation Theory, EDT) [16] que argumenta que la satisfacción de los clientes viene determinada por la impresión que éstos experimentan después de una compra como resultado de la desconformidad, positiva o negativa, entre las expectativas y los sentimientos derivados de su experiencia [17]. A partir de este concepto, nosotros definiremos la Satisfacción como la actitud general que el consumidor manifiesta por la experiencia acumulada de su comportamiento de compra y que en nuestro caso queda recogida en las bases de datos de las entidades financieras de manera que analizando la información de las mismas y conociendo el nivel de satisfacción para cada usuario, la entidad financiera podrá determinar los perfiles de usuarios de manera que sea posible corregir las desviaciones que se detecten y se potencien aquellas que sean rentables. [18], pone de manifiesto como la intención de continuar realizando transacciones financieras a través de Internet viene determinada por la satisfacción con las experiencias anteriores, el valor percibido del servicio y el programa de fidelización desarrollado por la empresa. Por tanto, cuando un usuario se encuentre satisfecho durante el desarrollo del servicio, se espera que su comportamiento a corto plazo sea congruente con su satisfacción.

2.. Métodos y Técnicas

A continuación, se describirán las técnicas usadas para poder realizar una correcta identificación de las variables relevantes que influyen en la confianza global del cliente así como del modelo que establecerá la relación.

2.1. Información Mutua

El concepto de Información Mutua (*Mutual Information*, MI), también conocido como entropía cruzada ha sido utilizado para resolver el problema de selección de variables en numerosas ocasiones [], mostrándose muy efectivo. Sea $X = \{\vec{x}_k\}$ e $Y = \{y_k\}$ para $k = 1..n$ entonces la MI entre X e Y puede ser definida como la cantidad de información que el grupo de variables X proporcionan sobre la variable de salida Y y se expresa como:

$$I(X, Y) = H(Y) - H(Y|X). \quad (1)$$

donde $H(Y)$ es la entropía de la variable Y , la cual mide la incertidumbre sobre Y . En el caso de tener variables continuas y, siguiendo la formulación de Shannon, puede definirse como:

$$H(Y) = - \int \mu_Y(y) \log \mu_Y(y) dy, \quad (2)$$

donde $\mu_Y(y)$ es la función de densidad marginal. Esta función puede definirse a su vez como la unión de la función de densidad de probabilidad $\mu_{X,Y}$ de X con Y , es decir:

$$\mu_Y(y) = \int \mu_{X,Y}(x, y) dx. \quad (3)$$

La entropía condicional $H(Y|X)$, que mide la incertidumbre de Y dado un X conocido, se define en el caso continuo como:

$$- \int \mu_X(x) \int \mu_Y(y|X = x) \log \mu_Y(y|X = x) dy dx. \quad (4)$$

La conclusión que se obtiene es que la información mutua $I(X, Y)$ es el grado de disminución de la incertidumbre sobre Y una vez que X es conocido. Gracias a las propiedades de la entropía y la información mutua, ésta última puede definirse también como:

$$I(X, Y) = H(X) + H(Y) - H(X|Y), \quad (5)$$

lo cual nos lleva a:

$$I(X, Y) = \int \mu_{X,Y}(x, y) \log \frac{\mu_{X,Y}(x, y)}{\mu_X(x)\mu_Y(y)} dx dy. \quad (6)$$

Esto nos indica que para calcular la información mutua entre dos grupos de variables pertenecientes a X , solo es necesario estimar la función de densidad de probabilidad conjunta entre X y Y . Aunque para realizar esta estimación se pueden emplear diversas técnicas como se describe en [19], la que se utiliza para la inicialización de los individuos se basa en el algoritmo de los K vecinos más cercanos (*K-Nearest Neighbors*, KNN) debido a que posee ciertas ventajas sobre los otros métodos basados en histogramas o en *kernels* [20].

Se ha utilizado la implementación disponible en [21] para calcular la información mutua. Esta implementación requiere la inicialización de un parámetro que define el número de vecinos más cercanos que se ha de considerar en la estimación. Tal y como se recomienda en [22] se ha fijado este valor a 6.

2.2. Test Delta

Otra técnica muy usada para identificar un subconjunto de variables representativas es el Test Delta [24]. Este test realiza una estimación del ruido existente entre pares de puntos de entrada/salida, por tanto, nos permitirá saber con cuanta precisión se puede modelar el conjunto de datos sin caer en el sobreentrenamiento. La aplicación al problema de selección de variables es bastante directa ya que el subconjunto de variables que proporcione un menor valor de ruido será el más recomendable usar para diseñar el modelo [23].

El Test Delta para un conjunto de muestras $X = \{\vec{x}_k\}$ e $Y = \{y_k\}$ para $k = 1 \dots n$ $(x_1, y_1), \dots, (x_N, y_N) \subset \mathbb{R}^d \times \mathbb{R}$, donde N es el número de muestras y d es la dimensionalidad de la entrada, puede ser definido como:

$$\delta_{N,k} = \frac{1}{2N} \sum_{i=1}^N (y_i - y_{nn[i,k]})^2 \quad (7)$$

donde $nn[i, k]$ es el índice del k -ésimo vecino más cercano a x_i . Aunque el criterio para medir la distancia puede variar, el más usado es la distancia Euclídea.

Sabiendo que $\delta_{N,1} \approx \sigma_e^2$, donde σ_e^2 es la varianza del ruido en la salida, $\delta_{N,1}$ puede utilizarse como una estimación del mejor error cuadrático medio que se puede obtener al diseñar un modelo que aproxime los datos sin sobreajuste.

El principal problema de este test es que su robustez y confianza está limitada por el número de elementos del conjunto de entrenamiento ya que converge al valor σ_e^2 al aumentar N [24,25]. Por tanto, si N no es lo suficientemente grande, la estimación puede no ser muy fiable.

2.3. Redes Neuronales de Funciones de Base Radial

Dentro del amplio abanico de tipos de redes neuronales, existe un tipo de red cuyo nombre completo es red neuronal de funciones de base radial (Radial Basis Function Neural Network, RBFNN) [26]. Una RBFNN se define matemáticamente como:

$$\mathcal{F}(\vec{x}_k; C, R, \Omega) = \sum_{i=1}^m \phi(\vec{x}_k; \vec{c}_i, r_i) \cdot \Omega_i \quad (8)$$

donde m es el número de neuronas en la capa oculta, $C = \{\vec{c}_1, \dots, \vec{c}_m\}$ son los centros de cada una de las RBF, $R = \{r_1, \dots, r_m\}$ son los valores de los radios de las RBF, $\Omega = \{\Omega_1, \dots, \Omega_m\}$ es el conjunto de los pesos correspondientes a cada una de las unidades de procesamiento y $\phi(\vec{x}_k; \vec{c}_i, r_i)$ representa a una RBF. Las RBFs más usadas para el problema de aproximación funcional son las Gaussianas debido a su caracter local que permite cubrir una región concreta del espacio de entrada [27–30]:

$$e^{-\frac{\|\vec{x}_k - \vec{c}_i\|^2}{r_i^2}} \quad (9)$$

Dentro del procedimiento para poder diseñar una red neuronal de RBFs es necesario determinar el valor de los parámetros de cada función además de la topología de la red, es decir, el número de neuronas que componen la red. Los parámetros a definir para cada RBF son la posición del centro y la amplitud para cada RBF. Una vez establecidos los valores y el número de RBF, la obtención del valor de los pesos puede plantearse como un sistema de ecuaciones lineal entre la salida de la función a aproximar y la matriz de activación de las RBF (P):

$$\vec{y} = P\Omega \quad (10)$$

donde P es una matriz de dimension $n \times m$ en la que P_{ki} representa el valor de la i -ésima función Gaussiana cuando se le proporciona como entrada el vector de entrada \vec{x}_k , siendo n igual al número de vectores de entrada. Hay muchas formas de resolver este sistema de ecuaciones. Entre las más usadas se encuentran la descomposición de Cholesky [31], la descomposición en valores singulares (*Singular Vale Decomposition*, SVD) [32] y el método de mínimos cuadrados ortogonales (*Orthogonal Least Squares*, OLS) [33].

Para dar un valor a los centros y los radios se diseñó el algoritmo OVI (Output Value-Based Initializer) [34,35], permitiendo dar un punto de partida adecuado, evitando caer en mínimos locales precipitadamente, para comenzar a realizar una búsqueda local que haga un ajuste fino de estos parámetros.

2.3.1. Diseño de las RBFNN: algoritmo OVI

El algoritmo OVI define una función de distorsión que es minimizada mediante un proceso de optimización que alterna el cálculo de las posiciones de los centros y los grados de activación para cada vector de entrada. Las posiciones de los centros y los grados de activación de cada neurona son actualizados en cada iteración del proceso de minimización y, una vez que estos valores no cambien significativamente, el algoritmo se detendrá. La función de distorsión es definida como:

$$J_{l,p}(A, C) = \sum_{k=1}^n \sum_{i=1}^m D_{ik}^2 a_{ik}^l |Y_k^p| \quad (11)$$

donde D_{ik} representa la distancia Euclídea desde un centro \vec{c}_i a un vector de entrada \vec{x}_k , a_{ik} es el grado en el que la neurona \vec{c}_i es activada por el vector de entrada \vec{x}_k , l es un parámetro que controla el grado en solapamiento de las neuronas, Y_k es la salida preprocesada del vector de entrada \vec{x}_k y p es un parámetro que permite modificar la influencia de la salida a la hora de calcular los grados de activación, haciéndola mayor o menor. La matriz de activación $A = [a_{ik}]$ está sujeta a las siguientes restricciones:

- $\sum a_{ik} = 1 \forall k = 1 \dots n$
- $0 < \sum a_{ik} < n \forall i = 1 \dots m$.

La primera restricción fuerza a cada vector de entrada a activar como mínimo a una neurona siendo 1 el valor máximo de activación. Gracias a la segunda, se puede garantizar que las neuronas serán activadas por al menos un vector de entrada. Esta restricción también nos permite mantener un solapamiento entre las neuronas permitiendo una interpolación más uniforme de la función. Las ecuaciones que permiten calcular las posiciones de las neuronas y sus grados de activación son:

$$a_{ik} = \left(\sum_{j=1}^m \left(\frac{D_{ik}}{D_{jk}} \right)^{\frac{2}{l-1}} \right)^{-1} \quad (12)$$

$$\vec{c}_i = \frac{\sum_{k=1}^n a_{ik}^l \vec{x}_k |Y_k^p|}{\sum_{k=1}^n a_{ik}^h |Y_k^p|} \quad (13)$$

donde D_{ik} es la distancia euclídea entre \vec{x}_k y \vec{c}_i , y Y_k es la salida preprocesada de la función a aproximar.

El algoritmo OVI, gracias a el cálculo de los grados de activación basándose en la salida preprocesada, es capaz de determinar un valor para el radio de cada neurona. Este valor es igual a la distancia del vector más lejano que activa a la neurona. Se puede calcular la longitud del radio haciéndola igual a la distancia Euclídea entre la posición de la neurona y la posición del vector de entrada con un valor de activación para esa neurona igual al valor del umbral:

$$r_i = \text{máx}\{ D_{ik} / a_{ik} > \vartheta_{overlap}, 1 \leq i \leq m, 1 \leq k \leq n \} \quad (14)$$

3.. Resultados

3.1. Base de datos

Los datos facilitados se derivan de la encuesta de satisfacción sobre Banca Electrónica que Caja Rural de Granada realiza a sus clientes cada dos ejercicios en la parte autenticada de su portal online denominado Ruralvía. Caja Rural de Granada es una Cooperativa de Crédito que forma parte del Grupo Caja Rural con una estructura de 203 oficinas (en las provincias de Granada, Málaga, Almería y Madrid), una red de 234 cajeros y una plantilla de 832 empleados y un resultado de 20,1 millones de euros al cierre del ejercicio 2010. La encuesta fue realizada entre los meses de septiembre y octubre de 2009. La participación en la encuesta era de carácter voluntario y se presentaba al usuario una vez que accedía a la parte autenticada de la web de Caja Rural de Granada. El tamaño muestral obtenido en la encuesta ascendió a 946 cuestionarios finalizados de usuarios únicos.

1. Oficina
2. Zona Geográfica
3. Margen ordinario del cliente para la entidad en el ejercicio 2010
4. Margen ordinario del cliente para la entidad en el ejercicio 2009
5. Saldo Medio de Pasivo Dentro de Balance del cliente en el ejercicio 2010
6. Saldo Medio de Pasivo Dentro de Balance del cliente en el ejercicio 2009
7. Saldo Medio de Pasivo Fuera de Balance del cliente en el ejercicio 2010
8. Saldo Medio de Pasivo Fuera de Balance del cliente en el ejercicio 2009
9. Saldo Medio de Activo del cliente en el ejercicio 2010
10. Saldo Medio de Activo del cliente en el ejercicio 2009
11. Número de productos contratados del cliente en el ejercicio 2010
12. Número de productos contratados del cliente en el ejercicio 2009

13. Vinculación por productos del cliente en el ejercicio 2010
14. Vinculación por productos del cliente en el ejercicio 2009
15. Vinculación por volumen de negocio del cliente en el ejercicio 2010
16. Vinculación por volumen de negocio del cliente en el ejercicio 2009
17. Vinculación por rentabilidad del cliente en el ejercicio 2010
18. Vinculación por rentabilidad del cliente en el ejercicio 2009
19. Nómina domiciliada
20. Pensión domiciliada
21. Tarjeta de débito contratada
22. Tarjeta de crédito contratada
23. Edad
24. Sexo
25. Teléfono móvil
26. Correo electrónico
27. Cliente autónomo
28. Código postal
29. Provincia
30. Meses de experiencia con Ruralvía
31. Número de operaciones de Ruralvía en el ejercicio 2010
32. Número de operaciones de Ruralvía en el ejercicio 2009
33. Importe de las operaciones de Ruralvía en el ejercicio 2010
34. Importe de las operaciones de Ruralvía en el ejercicio 2009

3.2. Selección de variables

Los resultados obtenidos tras aplicar los dos métodos se han representado gráficamente en la figura 1. Los valores obtenidos tras el cálculo ambos métodos han sido normalizados entre 0 y 1. En concreto, para el valor de Test delta, dado que "menor es mejor", una vez normalizados los resultados, se han invertido de modo que la métrica pasa a ser "mayor es mejor" igualándose con la información mutua.

En nuestro estudio para definir la satisfacción de los usuarios de Banca Electrónica, las variables que se han considerado determinantes para definir la satisfacción han sido: - Atendiendo al criterio de información mutua, las variables seleccionadas han sido el importe de las operaciones efectuadas a través de Banca Electrónica (33), el saldo de pasivo dentro de balance que mantiene el cliente con la entidad (5 y 6), la vinculación por rentabilidad para la entidad (18) y el número de operaciones efectuadas a través de Banca Electrónica (34). - Atendiendo al delta test, las variables seleccionadas han sido el sexo (24), la provincia (29), la domiciliación de la pensión (20) y la vinculación por rentabilidad (17 y 16).

Tras exponer los datos a los responsables de Desarrollo de Negocio de Caja Rural de Granada, se establecen ambos criterios de selección como válidos, si bien es cierto que el primero de ellos atiende en mayor medida a criterios de orden económico-financiero y el segundo de ellos a criterios socio-demográficos.

Desde la perspectiva de la gestión de una entidad financiera, será necesario por tanto conseguir vincular a los clientes para que realicen un mayor número y volumen de operaciones de Banca Electrónica, lo que supondrá una mejora de la rentabilidad para la entidad, y un incremento de la satisfacción de los clientes.

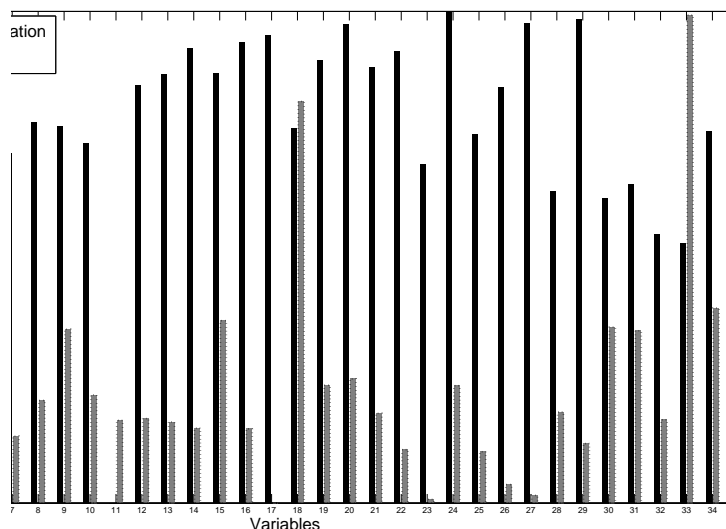


Figura 1: Clasificación de las variables usando el Test Delta (negro) y la información mutua (gris). Los valores están normalizados entre 0 y 1 y debe considerarse que el Test Delta ha sido invertido de modo que 1 es el mejor valor de Delta Test.

3.3. Regresión

En esta sección se van a realizar distintos diseños de RBFNN para ver la capacidad de aproximación de estos modelos de cara a predecir el grado de satisfacción global de un cliente. Para los experimentos se fija manualmente el número de neuronas que tendrá la red.

Siendo conscientes de la limitación en el número de datos y para evitar problemas causados por conjuntos de datos no balanceados, se realiza una validación cruzada, definiendo 4 subconjuntos de datos compuestos por 234 datos para test y el resto para entrenamiento (train).

A partir de los dos rankings obtenidos, se han ido construyendo RBFNN incrementando el número de variables: la primera red utiliza la variable más significativa del ranking, la segunda red utiliza las dos más significativas, etc. Los resultados de aproximación para los dos métodos de selección de variables se muestran en la Tabla 1.

Cuadro 1: Media y desviación estándar del Error Cuadrático Medio Normalizado (NRMSE) obtenido para las aproximaciones de los distintos conjuntos de entrenamiento y test usando los dos rankings de variables.

#variables	Train MI	Train DT	Test MI	Test DT
1	0.9927(0.0014)	0.9893(0.0046)	11.4261(23.86)	1.0206 (0.0173)
2	0.9900(0.0057)	0.9767(0.0054)	26.63(65.23)	23.977 (5.87)
3	0.9954(0.0084)	0.9735(0.0051)	1.0431(0.0423)	9.2236 (2.25)
4	0.9888(0.0056)	0.97382(0.0043)	1.5013(1.060)	8.4396 (18.25)
5	0.9918(0.0058)	0.96435(0.01604)	1.0536(0.0512)	1.066 (0.0990)

En los resultados puede apreciarse como, si bien el Test Delta permite alcanzar menores errores de aproximación con las variables elegidas, el modelo, a la hora de entrenarse cae en el sobreajuste siendo incapaz de generalizar. Por otra parte, cuando la información mutua es usada, los errores de entrenamiento son ligeramente superiores pero de cara a generalizar, el modelo se comporta mejor. Los resultados tan dispares son consecuencia del baraje de los datos ya que hay pocas instancias con valores de confianza pequeños.

4. Conclusiones

Durante la última década el sistema financiero español ha sufrido una profunda transformación que no ha conseguido reducir la importante tasa de bancarización que existe. Uno de los cambios de mayor importancia para el sector ha sido la revolución que ha supuesto la implantación de la Banca Electrónica entre los clientes. La Banca Electrónica cuenta en la actualidad con una penetración próxima al 50 % de los usuarios de Internet y al 20 % [36, 37] del total poblacional según diferentes fuentes, lo que revela la importante aceptación que ha supuesto para la sociedad este servicio. Esta importancia ha obligado a las entidades

a realizar importantes esfuerzos económicos para mantener y reforzar la satisfacción de sus usuarios en un canal de relevancia manifiesta y candidato a modificar el concepto tradicional de entidad financiera. La importancia de la Banca Electrónica viene motivada por el elevado número de ventajas que supone su uso en relación con un reducido número de inconvenientes. La integración de la Banca Electrónica en la estrategia multicanal de las entidades financieras proporcionará un incremento de las sinergias y en consecuencia una mayor calidad de servicio así como un mayor nivel de satisfacción. Esto permitirá relaciones rentables con los clientes siempre se descubran las necesidades de los mismos y se satisfagan eficientemente. La satisfacción es considerada por las entidades financieras una variable estratégica ya que define el nivel de fidelidad a través de diferentes dimensiones [?, ?, 38] y el resultado mejorado que la entidad pueda obtener en función del nivel de satisfacción de sus clientes [39, 40]. Por estos motivos el conocimiento de los perfiles de los usuarios de Banca Electrónica a partir de la información almacenada en las bases de datos de las entidades, permitirá seleccionar variables que posibiliten planificar estrategias de mejora para alcanzar los objetivos establecidos con anterioridad. Tras exponer los datos a los responsables de Desarrollo de Negocio de Caja Rural de Granada, se establecen ambos criterios de selección como válidos, si bien es cierto que la información mutua ha destacado criterios de orden económico-financiero y el test delta ha destacado criterios socio-demográficos. Desde la perspectiva de la gestión de una entidad financiera, será necesario por tanto conseguir vincular a los clientes para que realicen un mayor número y volumen de operaciones de Banca Electrónica, lo que supondrá una mejora de la rentabilidad para la entidad, y un incremento de la satisfacción de los clientes.

Agradecimientos Este trabajo de investigación ha sido financiado por los proyectos Spanish CICYT TEC2008-04920, Junta Andalucía P08-TIC-03674, P08-TIC03928 y PYR-2010-17 del CEI BioTIC GENIL (CEB09-0010) del MICINN.

REFERENCES

- [1] J. McCarthy. *Basic Marketing: a Managerial Approach*. Irwin, Homewood, Illinois, 1964.
- [2] W. Stanton. "Fundamentos de Marketing". 1970.
- [3] P. Kotler. *Marketing Management. Analysis, Planning and Control*. Prentice Hall, New York, USA, 1976.
- [4] L. Berry. "Relationship Marketing of Services - growing interest, emerging perspectives". *Journal of the Academy of Marketing Sciences*, vol. 23, pp. 236–245, 1995.
- [5] C. Grönroos. "The marketing strategy continuum: towards a marketing concept for the 1990s". *Management Decision*, vol. 29, pp. 9–9, 1991.
- [6] H. Karjaluoto, M. Mattila and T. Pentto. "Factors underlying attitude formation toward online banking in Finland". *International Journal of Bank Marketing*, vol. 20, pp. 261–272, 2002.
- [7] J. Torrent-Sellens, D. Castillo, P. Gabaldon, E. Ruiz and J. Sainz. *Hacia la banca multicanal*. ESIC., Madrid, Spain, 2010.
- [8] A. Asociación de Usuarios de Internet. "Informe de la A.U.I. sobre el uso de la Banca Electrónica en España". 2001.
- [9] F. Climent and A. Momparler. "La situación de la Banca On line en España". *Boletín Económico de ICE*, pp. 27–49, December 2006.
- [10] M. Berrocal. "Fidelización y Venta Cruzada". *Informe Caja Castilla La Mancha*, 2009.
- [11] B. for International Settlements. "Risk management principles for electronic banking". May 2001.
- [12] Fujitsu. "Soluciones para el Sector Bancario". *Informe Fujitsu*, 2009.
- [13] R. Cardozo. "An experimental study of consumer effort, expectations and satisfaction". *Journal of Marketing Research*, pp. 244–249, August 1965.
- [14] J. A. Howard and J. N. Sheth. *The Theory of Buyer Behavior*. Wiley, New York, USA, 1969.
- [15] B. Moliner. *La formación de la satisfacción/insatisfacción del consumidor y del comportamiento de queja: aplicación al ámbito de los restaurantes*. Universidad de Valencia, Valencia, Spain, 2004.
- [16] R. Oliver. "A cognitive model of the antecedents and consequences of satisfaction decisions". *The International Journal of culture, Tourism and Hospitality Research*, pp. 460–469, 1980.
- [17] D. Setó. "La influencia de la calidad de servicio, la imagen, la satisfacción y la confianza en la fidelidad del cliente". *Revista Española de Investigación de Marketing ESIC*, vol. 7, pp. 55–60, 2003.
- [18] A. Bhattacharjee. "An empirical analysis of the antecedents of electronic commerce service continuance". *Decision Support Systems*, pp. 201–214, 2001.
- [19] B.V. Bonnlander and A.S. Weigend. "Selecting input variables using mutual information and nonparametric density estimation". In *Proc. of the ISANN*, Taiwan, 2004.

- [20] A. Kraskov, H. Stögbauer and P. Grassberger. “Estimating mutual information”. *Physics Review*, June 2004.
- [21] H. Stögbauer, A. Kraskov, S. A. Astakhov and P. Grassberger. “Least-dependent-component analysis based on mutual information”. *Phys. Rev. E*, vol. 70, no. 6, pp. 066123, Dec 2004.
- [22] H. Stögbauer and A. Kraskov and S. A. Astakhov and P. Grassberger. “Least dependent component analysis based on mutual information”. *Physics Review*, December 2004.
- [23] A. Lendasse, F. Corona, J. Hao, N. Reyhani and M. Verleysen. “Determination of the Mahalanobis matrix using nonparametric noise estimations”. In *ESANN*, pp. 227–232, 2006.
- [24] H. Pi and C. Peterson. “Finding the Embedding Dimension and Variable Dependencies in Time Series”. *Neural Computation*, vol. 6, no. 3, pp. 509–520, 1994.
- [25] D. Evans. “The Gamma Test: Data derived estimates of noise for unknown smooth models using near neighbour asymptotics”, 2002.
- [26] D. S. Broomhead and D. Lowe. “Multivariate Functional Interpolation and Adaptive Networks”. *Complex Systems*, vol. 2, pp. 321–355, 1988.
- [27] A. G. Bors. “Introduction of the Radial Basis Function (RBF) networks”. *OnLine Symposium for Electronics Engineers*, vol. 1, pp. 1–7, February 2001.
- [28] I. Rojas, M. Anguita, A. Prieto and O. Valenzuela. “Analysis of the operators involved in the definition of the implication functions and in the fuzzy inference process”. *International Journal of Approximate Reasoning*, vol. 19, pp. 367–389, 1998.
- [29] T. Poggio and F. Girosi. “Networks for approximation and learning”. In *Proceedings of the IEEE*, volume 78, pp. 1481–1497, 1990.
- [30] J. Park and J. W. Sandberg. “Universal approximation using radial basis functions network”. *Neural Computation*, vol. 3, pp. 246–257, 1991.
- [31] H. Pomares, I. Rojas, J. Ortega, J. González and A. Prieto. “A Systematic Approach to a Self-Generating Fuzzy Rule-Table for Function Approximation”. *IEEE Trans. Syst., Man, Cyber. Part B*, vol. 30, no. 3, pp. 431–447, June 2000.
- [32] P. P. Kanjilal and D. Banerjee. “On the Application of Orthogonal Transformation for the Design and Analysis of Feedforward Networks”. *IEEE Trans. Neural Networks*, vol. 6, no. 5, pp. 1061–1070, 1995.
- [33] S. Chen, C. F. Cowan and P. M. Grant. “Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks”. *IEEE Trans. Neural Networks*, vol. 2, pp. 302–309, 1991.
- [34] A. Guillén, J. González, I. Rojas, H. Pomares, L. Herrera, O. Valenzuela and F. Rojas. “Output Value-Based Initialization For Radial Basis Function Neural Networks”. *Neural Processing Letters*, vol. DOI:10.1007/s11063-007-9039-8, June 2007.
- [35] A. Guillén, I. Rojas, J. González, H. Pomares, L. Herrera and A. Prieto. “Supervised RBFNN Centers and Radii Initialization for Function Approximation Problems”. In *International Joint Conference on Neural Networks, 2006. IJCNN '06*, pp. 5814–5819, July 2006.
- [36] INE. “Encuesta sobre Equipamiento y Uso de Tecnologías de la Información y Comunicación en los hogares”. 2010.
- [37] F. Orange. “eEspaña 2010: Informe anual sobre el desarrollo de la sociedad de la información en España”. pp. 108–115, 2010.
- [38] L. Methile and H. Nysveen. “Loyalty of On-Line Bank Customers”. *Journal of Information Technology*, vol. 14, pp. 375–386, 1999.
- [39] S. Athanassopoulus, S. Gounaris and V. Stathakopoulus. “Behavioural responses to customer satisfaction: an empirical study”. *European Journal of Marketing*, vol. 35, pp. 687–707, 2001.
- [40] P. Kotler and G. Armstrong. *Marketing: Edición para Latinoamérica*. Prentice Hall, Madrid, Spain, 2003.