



Industrial Data

ISSN: 1560-9146

iifi@unmsm.edu.pe

Universidad Nacional Mayor de San Marcos
Perú

Salinas Flores, Jesús

Patrones de morosidad para un producto crediticio usando la técnica de árbol de clasificación CART

Industrial Data, vol. 8, núm. 1, enero, 2005, pp. 29-36

Universidad Nacional Mayor de San Marcos

Lima, Perú

Disponible en: <http://www.redalyc.org/articulo.oa?id=81680106>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

PATRONES DE MOROSIDAD PARA UN PRODUCTO CREDITICIO USANDO LA TÉCNICA DE ÁRBOL DE CLASIFICACIÓN CART

Recepción: Febrero de 2005 / Aceptación: Junio 2005

⁽¹⁾ Jesús Salinas Flores

RESUMEN

El estudio tiene como objetivo encontrar un patrón de comportamiento de la morosidad a partir de la información obtenida al momento de solicitar un crédito para un producto crediticio y a su vez dar a conocer una nueva técnica estadística muy útil para este campo que es la técnica de árbol de clasificación CART la cual se aplica en situaciones donde se tiene un conjunto de datos de individuos en los que se han medido variables predictoras o independientes y una variable de clasificación o de criterio que define el grupo al que cada individuo pertenece; y se quiere encontrar un conjunto de reglas de decisión que permitan explicar la clasificación existente y utilizar estas reglas para poder clasificar a un nuevo individuo.

Palabras Clave: Riesgo crediticio. Árbol de clasificación. Algoritmo CART.

PATTERNS OF DILATORINESS FOR A CREDIT PRODUCT USING THE CART CLASSIFICATION TREE TECHNIQUE ABSTRACT

This study has the aim to find a pattern in delayed payments from the information obtained at the moment of requesting credit in a specific creditable product. At the same time, we show a very useful new statistical technique for this area, that is the classification tree (CART) that is applied in situations where we have independent predictor variables of classification or criterion that define the group to which every individual belongs. The paper also tries to find a set of decision rules that allow an explanation of the actual classification and the use of these rules to classify any new individual.

Key words: Credit risk. Classification tree. CART algorithm.

INTRODUCCIÓN

La evaluación de riesgos financieros de un banco o de una entidad crediticia es llevada a cabo por la gerencia de riesgo usando por lo general técnicas subjetivas o utilizando técnicas descriptivas univariadas/bivariadas o tomando como referencia resultados o modelos obtenidos en otras instituciones crediticias.

Cuando se otorgan préstamos o créditos a las personas naturales, se van a tener situaciones en las que se registran un conjunto de datos de "N" individuos o clientes clasificados en dos o más grupos ya definidos y en los que se han medido "p" variables predictoras o independientes y una variable de clasificación o de criterio que define el grupo al que cada individuo pertenece.

Para explicar la clasificación existente, las técnicas multivariadas tratan de encontrar una regla discriminante que clasifique correctamente a los individuos ya formados y posteriormente, aplicar esta regla a nuevos individuos. Entre las técnicas multivariadas disponibles se tienen al análisis de regresión múltiple, el análisis de regresión logística, el análisis discriminante lineal y cuadrático, etc.

Sin embargo, existen técnicas que permiten encontrar patrones de comportamiento basado en un conjunto de variables independientes, que son factibles de aplicar en el mercado peruano como son los denominados árboles de clasificación, los cuales son gráficos que ilustran reglas de decisión y que tienen como finalidad explicar la clasificación y predecir a los clientes que provienen de estos sistemas de clasificación, los cuales corresponden a una serie de reglas de decisión que predicen o clasifican observaciones futuras. Dentro de estos árboles de clasificación se tiene al algoritmo CART (*Classification and Regression Tree*).

El objetivo del presente trabajo es determinar las características o el reconocimiento de patrones de morosidad del producto crediticio de una entidad del sistema financiero nacional, como un método que discrimine mejor el proceso de otorgamiento de un crédito, utilizando el algoritmo de árbol de clasificación CART.

METODOLOGÍA DEL CART

Se tiene un conjunto de individuos u objetos O_1, \dots, O_n . Cada individuo O_i tiene asociado un vector x_i , siendo x_i una característica (variable) cuantitativa o cualitativa. La población $\hat{\Omega}$ está dividida en k grupos Π_1, \dots, Π_k tal como se presenta en el Cuadro 1.

(1) Magister en Ingeniería Industrial. Docente del Departamento de Estadística e Informática de la Universidad Nacional Agraria La Molina, UNALM. E-mail: jsalinas@lamolina.edu.pe

>>> Patrones de Morosidad para un Producto Crediticio usando la Técnica de Árbol de Clasificación CART

Cuadro 1. Datos para el algoritmo CART

Individuo	X_1	X_2	X_j	X_p	Grupo
O_1	x_{11}	x_{12}	...	x_{1j}	...	x_{1p}	1
O_2	x_{21}	x_{22}	...	x_{2j}	...	x_{2p}	1
.
.
O_i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{ip}	.
.
.
O_n	x_{n1}	x_{n2}	x_{nj}	x_{np}	K

Fuente: Elaboración propia, 2005.

Se quiere identificar secuencialmente, usando el algoritmo de árbol de clasificación CART, por orden de relevancia, las características X_1, \dots, X_p que mejor ayuden a identificar cada objeto como miembro del grupo Π_j al que pertenece.

Los árboles de decisión trabajan con dos tipos de variables: las variables criterio y las variables predictoras. Las variables criterio son aquellas cuyos resultados se desean predecir a partir de otras variables. También se denomina la variable dependiente. Las variables predictoras son las que predicen el patrón de la variable criterio. También se denominan variables independientes.

CART divide los datos en dos subconjuntos, de modo que los casos comprendidos dentro de cada uno de los subconjuntos sean más homogéneos que el subconjunto anterior. Se trata de un proceso recursivo, que se repite hasta alcanzar el criterio de homogeneidad o hasta llegar a otro criterio de detención. La misma variable predictora puede ser utilizada varias veces en distintos niveles del árbol.

Según Lewis (3), entre las ventajas del algoritmo CART se tienen:

- Es un método bastante flexible.
- CART no hace asunciones sobre la distribución de ningún tipo, ya sea de las variables predictoras como de la variable criterio.
- Las variables pueden ser de diferentes tipos como continuas, discretas, categóricas o nominales.
- CART no está afectado por valores extremos, colinealidad, heterocedasticidad que afecten los procedimientos establecidos. Los valores outliers pueden ser aislados en un nodo y no tienen ningún efecto en la división.
- CART puede detectar y revelar interacciones en un conjunto de datos.

- CART es invariante con relación a la transformación de las variables independientes. Es decir, las transformaciones no afectan el árbol producido.
- CART con el conjunto de variables predictoras o independientes, produce resultados usando sólo las variables más importantes.

No obstante, el mismo Lewis (3) menciona que el CART presenta algunos inconvenientes. Entre sus principales desventajas se tienen:

- Al ser un algoritmo binario, tiende a generar árboles de muchos niveles. Por ello, el árbol resultante puede que no presente los resultados de manera eficiente, sobre todo si la misma variable ha sido utilizada para la división de varios niveles consecutivos.
- CART no está basado en un modelo probabilístico. No existen intervalos de confianza asociados con las predicciones derivadas usando el algoritmo CART para clasificar a un conjunto de datos.

CART funciona eligiendo en cada nodo una división, de modo que cada nodo filial sea más puro que su nodo parental. Aquí, la pureza se refiere a los valores de la variable criterio. En un nodo completamente puro, todos los casos tienen el mismo valor para la variable criterio. CART mide la impureza de la división de un nodo dado definiendo una medida de impureza.

Para buscar las divisiones en los modelos CART se utilizan cuatro medidas de impureza distinta, las cuales dependen del tipo de variable criterio. Para las variables categóricas se puede elegir Gini, binario o (para criterios ordinales) binario ordenado. Para los criterios continuos, se utiliza el método de desviación mínimo cuadrática (LSD) en la búsqueda de la división. De todos ellos, el más usado es el índice de Gini.

El índice de Gini en el nodo t , $g(t)$, se define como:

$$g(t) = \sum_{j \neq i} p(j/t) p(i/t)$$

donde i y j son categorías de la variable criterio. La función del criterio Gini $\Phi(s, t)$ para la división s en el nodo t se define como

$$\Phi(s, t) = g(t) - p_L g(t_L) - p_R g(t_R)$$

donde p_L es la proporción de casos de t enviados al nodo filial de la izquierda, y p_R es la proporción enviada al nodo filial de la derecha. Se elige la división s que maximice el valor de $\Phi(s, t)$. Este valor, ponderado por la proporción de todos los casos del nodo t , es el valor del que se informa en el árbol como "mejora".

El algoritmo CART tiene los siguientes pasos:

1. Para llevar a cabo un análisis CART, comenzando por el nodo raíz $t=1$, buscar la división s^* , de entre todos los candidatos posibles S , que de lugar a la mayor reducción de la impureza:

$$\Phi(s^*, 1) = \max_{s \in S} \Phi(s, 1)$$

Luego dividir el nodo 1 ($t=1$) en dos nodos, $t=2$ y $t=3$, utilizando la división s^* .

2. Repetir el proceso de búsqueda de divisiones para uno de los nodos $t=2$ y $t=3$, y así sucesivamente.
3. Continuar con el proceso de desarrollo del árbol hasta alcanzar al menos una de las reglas de parada.

Se detiene el proceso de desarrollo del árbol cuando se cumple una de las diversas reglas de parada

disponibles. Un nodo no se dividirá si se cumplen alguna de las siguientes condiciones:

- Todos los casos de un nodo tienen valores idénticos en todos los predictores.
- El nodo se vuelve puro; esto es, todos sus casos tienen el mismo valor en la variable criterio.
- La profundidad del árbol ha alcanzado el valor máximo preestablecido.
- El número de casos que constituyen el nodo es menor que el tamaño mínimo preestablecido para un nodo parental.
- La división del nodo tiene como resultado un nodo filial cuyo número de casos es menor que el tamaño mínimo preestablecido para un nodo filial.

METODOLOGÍA DE LA INVESTIGACIÓN

El estudio es de tipo exploratorio y descriptivo. A través de una data histórica de clientes para un producto crediticio, del cual ya se conocen los clientes morosos y no morosos se pueden buscar patrones de comportamiento de morosidad y de no morosidad para incrementar la precisión al momento de evaluar el crédito para un nuevo cliente.

Para ello, se realizó un estudio en el período comprendido entre Agosto del 2001 y Junio del 2002 de todas las solicitudes aprobadas del producto crediticio. Se tomaron las variables almacenadas en la base de datos de la entidad bancaria y que fueron recogidas a través de los formularios de solicitud del crédito, y se seleccionaron para participar aquellas variables que tenían mayor poder predictivo. En el Cuadro 2 se muestra un listado y descripción de las mismas.

Cuadro 2. Descripción de las variables en estudio

Nombre de la variable	Descripción
MORA	Condición de moroso del cliente con la entidad bancaria. La entidad bancaria considera moroso a clientes con más de 90 días de atraso en una cuota. Esta variable (MORA) constituye la variable criterio.
EDAD	Edad (en años) del cliente.
SEXO	Sexo del cliente.
ESTCIVIL	Estado civil del cliente.
CARGAS	Carga familiar del cliente. Número de personas dependientes del cliente.
FONPAR	Si el cliente posee o no teléfono particular.
FONLAB	Si el cliente posee o no teléfono en su centro laboral.
CONAUTOV	Si presenta o no autovaluo al momento de solicitar el crédito.
ESAVAL	Si el cliente es aval o no de otro cliente de la entidad bancaria.
ANTLAB	La antigüedad laboral del cliente en años.
TIPREN	Si la renta del cliente es fija o variable.
UBIGEO	La ubicación geográfica (ciudad del país) desde donde se aprobó el crédito al cliente.

Fuente: Elaboración propia, 2005.

>>> *Patrones de Morosidad para un Producto Crediticio usando la Técnica de Árbol de Clasificación CART*

Se aplicó la siguiente metodología de investigación:

1. Seleccionar las variables predictoras más importantes.
2. Descripción univariada y bivariada de las variables descriptivas y de la variable criterio (morosidad).
3. Aplicación del algoritmo de árbol de clasificación CART.
4. Determinación de la pertinencia de las variables predictoras.
5. Obtención de las reglas de clasificación de la morosidad y de la no morosidad.

Para el procesamiento de la información se trabajó con la información de todos los clientes que obtuvieron el producto crediticio. El trabajo se realizó en dos etapas:

Etapla I: Descripción univariada y bivariada de las variables descriptivas y de la variable criterio, para lo cual se utilizó el software estadístico SPSS ver 10.0.

Etapla II: Aplicación del algoritmo de árbol de clasificación, para lo cual se utilizó el software CART 4.0 for Windows de la compañía Salford Systems; para generar los resultados de la clasificación usando como medida de impureza el índice de Gini con costos de mala clasificación iguales.

ANÁLISIS DESCRIPTIVO UNIVARIADO Y BIVARIADO

Antes de aplicar el algoritmo CART, se debe explorar previamente la información de la matriz de datos. El análisis univariado va a permitir describir a los clientes que solicitaron el crédito en función a cada una de las variables predictoras así como en función a la variable criterio de morosidad (MORA).

Se obtuvo que el producto crediticio durante el período de evaluación (Agosto 2001- Junio 2002) ha dado como resultado un porcentaje de 4,4% de morosidad, lo que constituye 32 clientes de un total de 724 clientes que solicitaron el crédito.

Al realizar el análisis univariado de las variables predictoras cuantitativas, se obtuvieron los siguientes resultados:

- La edad de los clientes que obtuvieron el crédito varía de 25 años hasta 66 años con una edad promedio de 42,31 años.
- La carga familiar de los clientes que obtuvieron el crédito varía de 0 a 4 con una carga familiar promedio de 1,45.
- La antigüedad de los clientes que obtuvieron el crédito varía de 1 año hasta 40 años con una antigüedad promedio de 9,9 años.

Al realizar el análisis univariado de las variables predictoras cualitativas, se obtuvieron los siguientes resultados:

- Los clientes que obtuvieron el crédito están constituidos en su mayoría (79,7%) por personas del sexo masculino
- Los clientes que obtuvieron el crédito están constituidos en su mayoría (63,5%) por personas casadas.
- Los clientes que recibieron el crédito, en su mayoría (82,3%) poseen teléfono particular.
- Los clientes que recibieron el crédito, en su mayoría (93,5%) poseen teléfono en su centro de trabajo. Esto indicaría que esta variable a pesar de su importancia no discriminaría la situación de moroso y no moroso.
- La mayoría de los clientes (85,2%) que obtuvieron el crédito han presentado el autoavalúo al momento de presentar la solicitud.
- Los clientes que obtuvieron el crédito, en su mayoría (73,9%) son avales de otros clientes de la entidad crediticia.
- Casi todos los clientes (95,6%) que obtuvieron el crédito tenían renta fija. Esto indicaría que esta variable a pesar de su importancia no discriminaría la situación de moroso y no moroso.
- La mayoría de clientes que obtuvieron el crédito provienen de Lima (38,3%).

Al realizar el análisis bivariado de las variables predictoras cuantitativas según la morosidad, se obtuvieron los siguientes resultados:

- No se puede afirmar que la edad influya sobre la morosidad o no morosidad. Se observa que la edad promedio de los clientes morosos y no morosos es similar (39,59 y 42,43 años).
- La carga familiar de los morosos es de 0,69 personas, mientras que de los no morosos es de 1,49 personas.
- Los clientes no morosos tienen una mayor antigüedad laboral que los morosos. Se observa que la antigüedad promedio de los morosos es de 6 años, mientras que de los no morosos es de 10 años.

Al realizar el análisis bivariado de las variables predictoras cualitativas según la morosidad, se obtuvieron los siguientes resultados:

- No se puede decir que los hombres o las mujeres son más morosos. Se observa que el porcentaje de morosidad es similar en ambos sexos (4,8% y 4,3%).
- Los solteros son más morosos que cualquier otra categoría del estado civil. Se observa que el porcentaje de morosidad de los solteros es el mayor de todos los estados civiles considerados (9,1%).
- Los clientes que no poseen teléfono particular

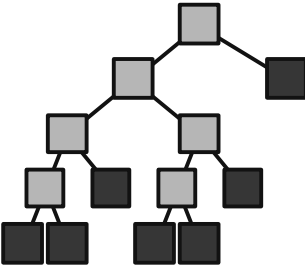


Figura 1. Árbol completo generado por el algoritmo CART

son más morosos que los que si lo poseen. Se observa que su porcentaje de morosidad es el mayor (8,6% vs. un 3,5%).

- No se puede decir que los clientes que poseen teléfono en su trabajo o los que no lo poseen son más morosos. Se observa que el porcentaje de morosidad es similar en ambos casos (4,3% y 4,4%).
- Los clientes que no presentaron autoavalúo al momento de solicitar el crédito son más morosos que los que sí lo presentaron. Se observa que su porcentaje de morosidad es el mayor (5,2% vs. un 0,0%).
- No se puede afirmar que los clientes con renta fija sean más morosos que los que poseen renta variable. Se observa que el porcentaje de morosidad es similar en ambos casos (4,3% y 6,3%).
- La morosidad está muy relacionada con el ubigeo donde se otorgó el crédito. En el caso de los morosos Junín posee la tasa más elevada de morosidad (25,0%).

ANÁLISIS DEL ÁRBOL GENERADO POR EL ALGORITMO CART

Usando el software CART y usando la variable MORA como variable criterio y a las demás variables como

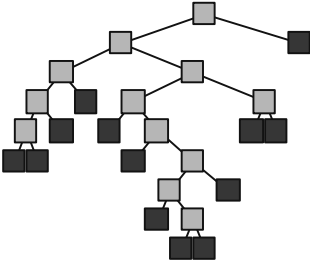


Figura 2. Árbol que describe comportamiento de morosidad

variables predictoras se obtiene el árbol de clasificación presentado en el Figura 1.

Antes de empezar a analizar el árbol de clasificación obtenido, se debe evaluar la pertinencia o no de cada variable predictora. El Cuadro 3 proporciona la importancia de cada una de las variables predictoras. Un score es asignado a cada variable, y está basado en la mejora que hace en la división del árbol. El score da un reconocimiento a las variables cuya importancia es marcada. Según se observa en este cuadro, las variables FONPAR (si el cliente posee o no teléfono particular), SEXO (sexo del cliente), TIPREN (tipo de renta del cliente) y FONLAB (si el cliente posee o no teléfono en su centro laboral) carecen de importancia (SCORE 0,00) como variables que puedan discriminar la morosidad o no morosidad de un cliente que fue sujeto a crédito.

Para el análisis posterior estas cuatro variables serán descartadas como variables predictoras y se procederá a realizar el árbol de clasificación usando las demás variables.

En la Figura 2 se aprecia el árbol conseguido para poder describir el comportamiento de los morosos y de los no morosos.

Cuadro 3. Importancia de cada variable predictora

Variable	Score	
UBIGEO	100,00	
ANTLAB	52,07	
CARGAS	47,69	
CONAUTOV	44,23	
ESTCIVIL	37,97	
EDAD	36,08	
ESAVAL	0,03	
FONPAR	0,00	
SEXO	0,00	
TIPREN	0,00	
FONLAB	0,00	

Fuente: Elaboración propia, 2005.

>>> Patrones de Morosidad para un Producto Crediticio usando la Técnica de Árbol de Clasificación CART

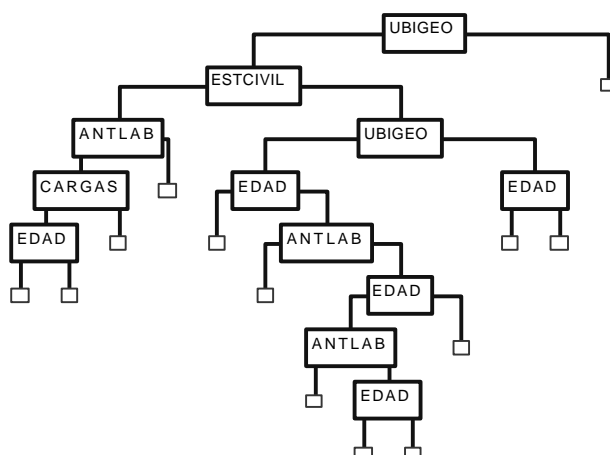


Figura 3. Árbol de clasificación con las variables predictoras que generaron el árbol

De la Figura 3 se pueden obtener los siguientes resultados:

- Es un árbol con 13 nodos terminales, mostrados de un color diferente (rectángulo rojo pequeño) de los nodos parentales (rectángulo azul grande).
- La primera variable predictor que va a discriminar a los morosos y los no morosos es el ubigeo o ciudad donde el cliente solicitó el crédito (UBIGEO).
- Las demás variables predictoras que van a discriminar a los morosos y no morosos son: el estado civil (ESTCIVIL), la antigüedad laboral (ANTLAB), la carga familiar (CARGAS) y la edad del cliente (EDAD).

IMPORTANCIA DE LAS VARIABLES PARA EL PRODUCTO CREDITICIO

El Cuadro 4 proporciona la importancia de todas las variables predictoras. Un score es asignado a cada variable, y está basado en la mejora que hace en la división del árbol. El score da un reconocimiento a las variables cuya importancia es marcada.

Según los resultados del árbol de clasificación las variables con los scores más altos han participado en la formación del árbol, esto es, son las variables que discriminan el árbol. Estas variables son: el ubigeo, la antigüedad laboral, la edad, la carga familiar y el estado civil. Las variables si presenta o no autovaluo, y si es aval no han participado en la formación del árbol.

RESULTADOS

Reglas de clasificación para los morosos
Las principales características o el patrón que poseen los morosos se resumen en lo siguiente:

- **Nodo 2:** Que el estado civil del cliente sea soltero y que el crédito lo haya obtenido en Arequipa, Callao, Junín, La Libertad, Lima o Tacna y la antigüedad laboral sea menor a 11 años y que la carga familiar sea menor de 2 hijos y que la edad del cliente sea mayor de 28 años. (representa el 68,75% de los clientes morosos); ó

Cuadro 4. Importancia de cada variable

Variable	Score	
UBIGEO	100,00	
EDAD	68,47	
ANTLAB	40,15	
ESTCIVIL	37,98	
CARGAS	15,99	
CONAUTOV	0,00	
ESAVAL	0,00	

Fuente: Elaboración propia, 2005.

Cuadro 5. Clasificación errónea

Clase	Número de casos	Número de casos mal clasificados	% de clasificación errónea	Número de casos correctamente clasificados	% de clasificación correcta
Moroso	32	2	6,25	30	93,75
No Moroso	692	94	13,58	598	86,42
Total	724	96	13,26	628	86,74

Fuente: Elaboración propia, 2005.

- **Nodo 7:** Que el crédito lo haya obtenido en Arequipa, Junín, La Libertad o Tacna y que el estado civil del cliente sea casado, divorciado o viudo y que la edad del cliente sea mayor a 42 años pero menor a 54 años y que la antigüedad laboral sea mayor a 6 años pero menor de 8 años (representa el 22,73% de los clientes morosos).

Reglas de clasificación para los no morosos
Las principales características o el patrón que poseen los no morosos se resumen en lo siguiente:

- **Nodo 13:** Que el lugar donde obtuvo el crédito sea Ancash, Apurímac, Ayacucho, Cajamarca, Cuzco, Huanuco, Ica, Lambayeque, Loreto, Moquegua, Piura, Puno, San Martín, Tumbes ó Ucayali (representa el 38,73 % de los clientes no morosos); ó
- **Nodo 12:** Que el lugar donde obtuvo el crédito sea Callao o Lima y que el estado civil sea casado, divorciado o viudo y que la edad sea mayor de 29 años (representa el 25,32% de los clientes no morosos).

Validación de los resultados

Para poder evaluar la capacidad del algoritmo CART a la hora de predecir la morosidad o no morosidad se

puede examinar el Cuadro 5 que compara el tipo asignado mediante el árbol con la morosidad/no morosidad realmente registrada. Se observa que la proporción de casos clasificados de manera incorrecta es bastante baja: en total 2 de 32 morosos y 94 de 692 no morosos. Esto da como resultado un porcentaje de clasificación correcta del 86,74% de todos los clientes que solicitaron el crédito, lo que indica que el algoritmo CART obtenido tiene poder predictivo.

En el Cuadro 6 se presentan algunos casos de clientes que solicitan el crédito y la predicción de su situación (moroso o no moroso) según el algoritmo CART desarrollado.

Medidas correctivas y control de mejoras
Los resultados obtenidos con el algoritmo CART parten de la premisa de que las solicitudes de crédito ya han pasado una evaluación estándar. La metodología sólo funciona para este tipo de créditos y en general para este producto crediticio.

El poder predictivo de los resultados del algoritmo CART supone que el desempeño de los créditos futuros tendrá las mismas características o condicio-

Cuadro 6. Simulación de predicciones

Edad	Estado civil	Carga familiar	Antigüedad laboral	Ubigeo	Nodo	Predicción según CART
33	Casado	2	5	Lima	12	No Moroso
44	Divorciado	1	11	Lima	12	No Moroso
32	Casado	1	5	Lima	12	No Moroso
43	Soltero	1	7	Callao	2	Moroso
37	Soltero	0	10	Lima	2	Moroso
36	Soltero	1	2	Ucayali	13	No Moroso
34	Soltero	1	2	Lima	2	Moroso
33	Soltero	0	4	Lima	2	Moroso
36	Soltero	0	7	Tacna	2	Moroso
36	Soltero	0	9	Lima	2	Moroso
54	Casado	0	29	Lambayeque	13	No Moroso
44	Casado	3	7	Tacna	7	Moroso

Fuente: Elaboración propia, 2005.

>>> *Patrones de Morosidad para un Producto Crediticio usando la Técnica de Árbol de Clasificación CART*

nes que los créditos pasados con los que se obtuvieron los resultados del algoritmo CART. Si las condiciones del producto crediticio cambia, el futuro no es como el pasado -como a menudo sucede a las instituciones financieras que desarrollan productos y nichos nuevos, enfrentan la competencia, u operan en mercados cambiantes- entonces los resultados del algoritmo CART no funcionarán bien, y en todo caso se deberá aplicar un nuevo algoritmo CART.

CONCLUSIONES

Usando el algoritmo CART se ha conseguido detectar las variables que más influyen sobre la morosidad/no morosidad del producto crediticio y se las aisló de las variables que no influyen. Estas variables son: el ubigeo, la antigüedad laboral, la edad, la carga familiar y el estado civil.

Usando el algoritmo CART se detectaron patrones diferentes para los morosos y los no morosos. Para los morosos el principal patrón detectado es que el estado civil del cliente sea soltero y que el crédito lo haya obtenido en Arequipa, Callao, Junín, La Libertad, Lima o Tacna y la antigüedad laboral sea menor a 11 años y que la carga familiar sea menor de 2 hijos y que la edad del cliente sea mayor de 28 años. Este patrón caracteriza al 68,75% de los clientes morosos. Para los no morosos el principal patrón detectado es que el lugar donde obtuvo el crédito sea Ancash, Apurímac, Ayacucho, Cajamarca, Cuzco, Huanuco, Ica, Lambayeque, Loreto, Moquegua, Piura, Puno, San Martín, Tumbes ó Ucayali.

Usando el algoritmo CART se ha automatizado el proceso crédito, discriminando con gran precisión si el cliente es sujeto a crédito. Se obtuvo un 93,75% de clasificación correcta para los morosos y un 86,42% de clasificación correcta para los no morosos. En general se logró un 86,74% de clasificación correcta. Además bajo este análisis de sistema de información la banca puede tener proyecciones más

seguras y a tiempo real de quienes son sus mejores clientes crediticios.

RECOMENDACIONES

Se pueden comparar los resultados obtenidos con el algoritmo CART con los resultados obtenidos con otros algoritmos de árboles de decisión entre los cuales se mencionan el CHAID (detector automático de interacciones de chi-cuadrado), CHAID exhaustivo y el algoritmo QUEST.

El algoritmo CART constituye una técnica para explorar datos que está limitado por la falta de un modelo probabilístico, por lo que sus resultados deberían ser confirmados con técnicas estadísticas multivariadas.

La capacidad de pronóstico aumenta con el número de características disponibles (variables independientes o explicativas) que se usen del formulario de solicitud del crédito.

BIBLIOGRAFIA

1. Bonson, E. (1999). *Tecnologías inteligentes para la gestión empresarial*. 1er. Ed. Editorial ra-ma. España.
2. Breiman, L., Friedman, R. Olshen, A. y Stone, C. (1984). *Classification and regression trees*. Belmont. 1er. Ed. Calif: Wadsworth. USA.
3. Lewis, R. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis*. Annual Meeting of the Society for Academy Emergency Medicine. San Francisco, California. USA.
4. Salford Systems. (2000). *Cart for Windows. User's guide*. A Salford systems implementation of the original CART program. 1er. Edic. USA.