GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

2014 - 2015

UNIVERSIDAD DE CANTABRIA

TRABAJO FIN DE GRADO

CREDIT SCORING

Autor: Pablo Sánchez Bilbao

Director: José Luis Gallego Gómez

Santander, Septiembre 2015

ÍNDICE

Resumen y palabras clave …….…….. Página 2

Abstract and key words …… Página 3

1. Introducción …….…….…….. Página 4-5

2. La teoría del *credit scoring* …….…….…... Página 5-10

2.1. Definición y objetivos …….….. Página 5

2.2. Variables y datos ……...…..... Página 5-6

2.3. Modelos estadísticos …….…….. Página 6-7

2.3.1. Análisis de regresión …… Página 6

2.3.2. Modelos de probabilidad no lineal logit y probit…….…… Página 7

2.4. Implementación de un sistema de *credit scoring* ……. Página 8-9

2.5. Software …….….. Página 9-10

3. La práctica del *credit scoring* …….…… Página 10-13

3.1. Banco Santander …….…….. Página 10-13

4. Aplicación práctica …….……... Página 13-22

4.1. Análisis exploratorio ……...…….. Página 13-18

4.2. Estimación del modelo…….……. Página 18- 19

4.3. Evaluación del modelo …… Página 20-21

4.4 Uso del modelo …… Página 21-22

5. Conclusiones ……... Página 23

6. Bibliografía ……...……... Página 24

Agradecimientos ……...…….. Página 25

RESUMEN

El *credit scoring* es un sistema de modelos de decisión a través del cual se calcula la probabilidad de que un sujeto sea capaz de devolver o no un crédito comercial.
Este trabajo realiza una simulación de este modelo con la intención de familiarizarse con el funcionamiento del mismo, así como con el marco teórico que este conlleva.

El uso del *credit scoring* da respuesta a un problema generalizado en las entidades financieras como es la correcta estimación y valoración del riesgo del cliente o, en este caso, del solicitante del crédito. Para realizar eficazmente esta estimación es necesario un análisis previo de los datos tanto personales como económicos del cliente así como de las características del crédito solicitado.
Los métodos clásicos estadísticos serán fundamentales para el desarrollo de este modelo, siendo los más utilizados el análisis de regresión y los modelos no lineales logit y probit.

La implementación de este sistema en las entidades financieras supone un importante cambio, puesto que los procesos de análisis de riesgo crediticio han pasado de realizarse manualmente a automatizarse. Esto conlleva una significativa reducción en tiempo y costes pero también importantes cambios en los hábitos de los trabajadores.

Los resultados obtenidos por este modelo facilitan a la entidad el trabajo, ya que se elimina el factor de subjetividad en las valoraciones y proporciona soluciones rápidas en la valoración del riesgo crediticio.

PALABRAS CLAVE

Credit scoring, riesgo, paneles de control, modelos estadísticos, software.

ABSTRACT

Credit Scoring is a decision-making model system through which the probability that a subject is able to repay a trade credit or not shall be calculated. This paper performs a simulation of this model intended to become acquainted with how the operation works, as well as the theoretical framework that this entails.

Use of credit scoring responds to a widespread problem in the financial institutions and their accurate estimation of customer´s credit risk profile. To effectively perform this estimate is necessary a prior analysis of personal customer data as well as economic characteristics of the credit requested.
Classical statistical methods will be fundamental for the development of this model. Regression analysis and nonlinear models logit and probit are most commonly used.

The implementation of this system in financial institutions represents an important change since the processes of credit risk analysis have been carried out manually to be automated. This leads to a significant reduction in time and costs but also important changes in the behaviour of workers.

The results obtained by this model make work in the institution easier due its subjectivity in assessments is eliminated and provides fast solutions in the valuation of credit risk.

KEY WORDS

Credit scoring, risk, control panels, statistical models, software.

1. INTRODUCCIÓN

Tras el estallido de la crisis financiera de 2008, la debilidad en las entidades financieras ha permanecido latente en estos últimos tiempos. Restricciones al crédito y problemas de liquidez han sido algunas de las consecuencias causadas por una mala gestión de riesgos por parte de estas entidades. Estas consecuencias las seguimos sufriendo a día de hoy; por ello, los sistemas que faciliten un uso eficiente de la gestión de riesgos o representen mecanismos de mayor control serán necesarios para que estas situaciones no se vuelvan a repetir.

Los acuerdos de Basilea II, publicados en 2004, ya recomendaron la necesidad de mejorar la calidad de los créditos con la medición de la probabilidad de incumplimiento (*probability of default*). Una de estas herramientas para este propósito es el *credit scoring*, que se puede definir como un sistema de modelos de decisión que permite a una entidad financiera calcular la probabilidad de que un cliente no devuelva un crédito. De esta manera la entidad podrá saber en cada caso qué riesgo está asumiendo al conceder ese crédito y podrá valorar hasta dónde estará dispuesta a llegar.

Este trabajo de fin de grado está centrado en el estudio de este sistema, siendo sus dos principales objetivos la descripción del marco teórico del *credit scoring* y de una metodología para su implementación práctica.

El marco teórico se basa principalmente en los libros de texto de Anderson (2007) y Thomas, Edelman y Crook (2002), que resumen una exhaustiva literatura sobre *credit scoring* y que son referencias bibliográficas básicas en los cursos de postgrado sobre esta materia. También se ha estudiado la documentación interna sobre *credit scoring* de dos entidades financieras como son el Banco de Santander y Liberbank. Se han mantenido varias reuniones con los directores de dos oficinas bancarias que me han facilitado el entendimiento de este sistema, ayudándome a comprender las ventajas que su aplicación supone y los cambios que se han efectuado en su día en cada una de las oficinas.

En cuanto a la metodología econométrica que se sigue en la implementación práctica del *credit scoring*, ha sido necesario el estudio de modelos econométricos avanzados que no se estudian en la titulación de GADE. En este sentido, ha sido de ayuda los materiales de la asignatura de Econometría Aplicada de GE.

Durante las reuniones con los directores de las oficinas del Banco de Santander y de Liberbank, se valoró la posibilidad de analizar una muestra de clientes que hubieran solicitado un crédito en estas oficinas. Aunque mostraron buena disposición a colaborar, las dificultades estaban en la naturaleza confidencial de los datos bancarios y en la imposibilidad de publicarlos para permitir reproducir los resultados. Por esta razón, se decidió analizar un conjunto de datos público que se estudia en numerosos estudios sobre aprendizaje de máquina: los datos de crédito alemán del profesor Hofmann de la Universidad de Hamburgo.

El trabajo se estructura, por tanto, en dos grandes bloques, la teoría del *credit scoring* y la implementación práctica del mismo. Dentro del primer bloque, primero se define y concreta el objetivo de este sistema. Después, se detalla la información socioeconómica que se debe recoger del cliente, así como las características del crédito solicitado. A continuación, se describen los modelos estadísticos más frecuentemente utilizados en el sistema de credit *scoring,* como son el análisis de regresión y los modelos de probabilidad no lineal probit y logit. Finalmente, se estudia el proceso de implementación de un sistema de *credit scoring* en una entidad financiera, señalando los principales factores que se deben tener en cuenta y el software más utilizado por las entidades. En

el bloque práctico, primero se describe el funcionamiento de los modelos de decisión en la entidad financiera Banco Santander y, después, se realizará la elaboración de un modelo de *credit scoring* usando los datos del crédito alemán. El trabajo finaliza con las conclusiones y extensiones.

2. LA TEORÍA DEL *CREDIT SCORING*

2.1. DEFINICIÓN Y OBJETIVOS

El *credit scoring* es un sistema de modelos de decisión mediante el cual se calcula la probabilidad de que un sujeto sea capaz de devolver o no un crédito comercial. Este sistema se basa en el uso de modelos estadísticos a través de los cuales se transformaran los datos previamente obtenidos del solicitante en medidas numéricas que servirán de apoyo en la toma de decisiones. De esta manera, la entidad emisora del crédito podrá decidir si concederlo o, en caso contrario, determinar que el sujeto no cumple los requisitos necesarios.

En la actualidad la gran mayoría de las entidades financieras utilizan estos sistemas de *credit scoring* tanto por su rapidez como por su fiabilidad. También presenta determinadas desventajas, como la gran dependencia que estos sistemas tienen de los datos.

Para que el sistema de *credit scoring* pueda realizar una correcta valoración del prestatario será necesario el análisis de las características del mismo y también de las características del crédito solicitado. El solicitante del crédito podrá ser una persona física o una persona jurídica, siendo diferente el análisis de datos en este segundo caso.

2.2. VARIABLES Y DATOS

Como se ha comentado, para que este sistema pueda emitir una adecuada valoración del cliente se debe realizar un análisis eficiente de sus datos. En primer lugar, los datos que recogerá la entidad serán de carácter más general, de esta manera se podrá realizar una primera valoración del sujeto. Estos datos dependerán del tipo de solicitante.

- *Trabajadores por cuenta ajena.* Documentación requerida:
  - Declaración de bienes de personas físicas
  - I.R.P.F
  - Nóminas
  - Títulos de propiedad
  - Capitulaciones matrimoniales en caso de separación de bienes
  - Tres últimos recibos del pago de deudas bancarias, en caso de que existiesen.

- *Pymes y autónomos.* Documentación requerida:
  - Justificantes de los tres últimos pagos a la Seguridad social
  - I.R.P.F ; IVA
  - Títulos de propiedad
  - Contratos y documentos que acrediten otros ingresos y fincabilidad
  - Tres últimos recibos del pago de deudas bancarias, en caso de que existiesen

- *Personas jurídicas.* Documentación requerida:
  - Justificantes de los tres últimos pagos a las Seguridad social

- Títulos de propiedad
- Contratos y documentos varios
- Tres últimos recibos del pago de deudas bancarias, en caso de que existiesen
- Impuesto de sociedades, IVA último año, IVA trimestrales

Toda la información recogida de los sujetos se debe de comparar con informes de otros terceros, antecedentes bancarios y, en el caso de personas jurídicas, consultas al registro de propiedad y la CIRBE (extracto del Banco de España). Estos datos no solo serán utilizados en la valoración del *credit scoring,* sino también en otros fines de interés para la entidad financiera (labores comerciales, etc.). De esta manera, la entidad elaborará una valoración primaria del solicitante que servirá de apoyo para decisiones futuras. Sin embargo, los datos que supondrán una mayor ponderación para la calificación crediticia serán bastante más concretos. Algunos ejemplos son: sexo, estado civil, nivel de estudios, nacionalidad, número de años residiendo en su domicilio, etc. Estas variables serán enumeradas y analizas mas adelante.

Otros datos que requieren ser analizados son los referentes al crédito solicitado. En particular, la entidad debe tener en cuenta principalmente tanto la cantidad total del crédito como la duración del mismo. Será de igual importancia el destino de ese préstamo a la hora de la estimación de la calificación crediticia, por lo que la puntuación asignada variará en función del propósito:

- Vehículo nuevo o vehículo de segunda mano
- Vacaciones
- Viviendas
- Ceremonias (bodas, bautizos…)
- Estudios
- Otros

2.3. MODELOS ESTADÍSTICOS

En este apartado se desarrollarán los modelos estadísticos clásicos utilizados a la hora de realizar el *credit scoring*: el análisis de regresión y los modelos de probabilidad no lineal logit y probit.

2.3.1. Análisis de regresión

En términos generales, el análisis de regresión trata sobre el estudio de una variable dependiente Y_i respecto de una o varias variables explicativas X_{ji}. La ecuación del modelo de regresión lineal general es:

$$Y_i = x'_i\beta + u_i, \quad i=1, 2,\ldots, n,$$

donde $x'_i\beta = \beta_1 + \beta_2 X_{2i} + ___ + \beta_k X_{ki}$, β_1 el término constante y β_j es la pendiente de la variable X_{ji} que mide la respuesta de la variable dependiente ante un cambio unitario de la variable explicativa, n es el número de datos, y u_i es el término de error aleatorio.

Cuando la variable dependiente es binaria, cada observación Y_i es una variable aleatoria Bernoulli con probabilidad de éxito $x'_i\beta$. La función de probabilidad de Y_i puede escribirse como:

$$P(Y_i) = (x'_i\beta)^{Y_i} (1 - x'_i\beta)^{1-Y_i}$$

y sus dos primeros momentos son:

$$E(Y_i) = x'_i\beta \qquad y \qquad V(Y_i) = x'_i\beta\,(1 - x'_i\beta)$$

El modelo de regresión se convierte en este caso, en un modelo probabilidad lineal (MPL), donde la esperanza matemática de la variable dependiente, $E(Y_i) = x'_i\beta$ es la probabilidad de éxito, que es una combinación lineal de los parámetros $\beta_1,\ldots,\beta_k$. Aunque el MPL puede estimarse usando mínimos cuadrados ordinarios, se presentan una serie de inconvenientes que cuestionan su uso:

- El término de error no sigue una distribución normal sino de tipo Bernoulli.
- El término de error presenta heterocedasticidad, la varianza no es constante.
- Algunas probabilidades estimadas pueden ser inadmisibles, negativas o mayores que la unidad.
- El R^2 suele tomar valores muy pequeñas porque la variable dependiente indica elecciones y los valores ajustados son las probabilidades de las mismas, es decir, el R^2 se basa en la correlación de una variable binaria y una variable continua.
- Los efectos marginales son constantes, es decir, los cambios de $P(Y_i)$ ante un cambio de X_{ji} no dependen de esta variable.

2.3.2. Modelos de probabilidad no lineal logit y probit

Teniendo en cuenta los inconvenientes anteriormente mencionados del modelo de probabilidad lineal, no deberían impedirnos apreciar la proposición fundamental del mismo: el componente sistemático $x'_i\beta$ es un indicador de la probabilidad de éxito. Esta proposición nos sugiere una manera de mejorar el MPL: expresar la probabilidad de éxito como una función no lineal de $x'_i\beta$,

$$\Pr(Y_i = 1) = F(x'_i\beta)$$

Esta función se corresponde con el modelo de probabilidad no lineal (MPNL) y debería cumplir las siguientes propiedades:

- Estar acotada entre 0 y 1, $0 \leq F(x'_i\beta) \leq 1$
- Ser monótona no decreciente, $x'_i\beta < x'_j\beta \rightarrow F(x'_i\beta) \leq F(x'_j\beta)$
- Tener límites en $\pm\infty$
- Ser continua en cualquier punto $x'_i\beta$

Estas cuatro propiedades que debe cumplir la función son las que definen a la función de distribución acumulada (FDA) de cualquier variable aleatoria continua. Por tanto, se pueden formular diferentes MPNL usando diferentes FDA. Los dos modelos de probabilidad no lineal más comunes son el modelo probit y el modelo logit.

1) *Modelo probit o normit* (la función es una FDA de una distribución normal estándar)

$$\Pr(Y_i = 1) = \int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)\,dz$$

2) *Modelo logit* (La función es la FDA de una distribución logística)

$$\Pr(Y_i = 1) = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}$$

2.4. IMPLEMENTACIÓN DE UN SISTEMA DE *CREDIT SCORING*

La aplicación de un modelo *credit scoring* conlleva una significativa reducción de los tiempos de actuación y decisión a la hora de evaluar y estimar el riesgo del crédito. De esta manera la intervención humana se simplifica, siendo de mayor importancia la automatización de los procesos. Sin embargo, la implementación y desarrollo de este sistema supone un coste elevado para la entidad, tanto por la formación y necesidad de un equipo especializado, como por el coste de desarrollo del *scoring* y las herramientas tecnológicas que este requiere.
Cuando se toma la decisión de implantar un sistema de *scoring* en una empresa o entidad, se deberá de tener en cuenta una serie de aspectos fundamentales para su correcto funcionamiento (figura 2.4.1).

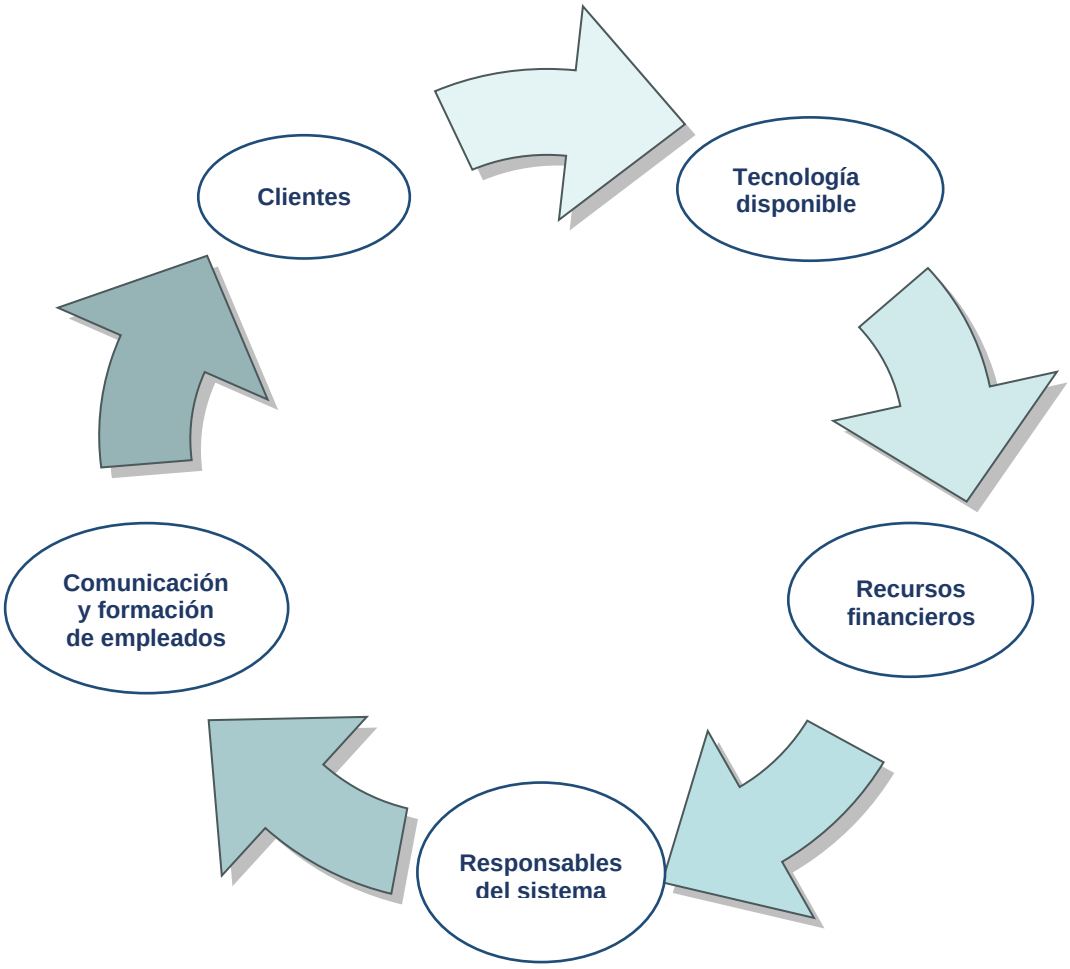

Figura 2.4.1: Factores a tener en cuenta en un proceso de implementación

- En primer lugar, la tecnología de la que dispone la entidad. Si la entidad ya dispone de paneles de control o equipos informáticos especializados, la inversión a realizar será menor. Se debe tener en cuenta que el coste de esta tecnología es muy elevado, por lo que la entidad podrá elegir entre dos sistemas: uno de coste bajo pero de respuesta más lenta llamado *score sheet*, y otro con un coste más elevado, pero a su vez de respuesta más rápida, con procesos automatizados, llamado *scorecard*.

- Los recursos financieros que se esté dispuesto a invertir en la implementación del *scoring* es otro factor importante que deberá de plantearse la empresa,

puesto que podemos encontrarnos con sistemas con precios muy dispares, como ya se ha comentado anteriormente. En la actualidad la gran mayoría de entidades bancarias optan por un sistema automático a pesar de su elevado coste.

- La responsabilidad de este sistema recaerá en el trabajador encargado de la instalación y el mantenimiento que el sistema requiere. En ocasiones estas tareas se delegan en negocios externos, empresas subcontratadas, como es el caso de las consultorías financieras.

- La comunicación a los empleados y su correspondiente formación, ya que esto supondrá un cambio en su hábito de trabajo. Uno de los puntos críticos a la hora de tomar la decisión de implantar un sistema de *scoring*, es el modo de comunicarlo a los empleados, buscando la forma más adecuada posible y eliminando cualquier inseguridad que surja, para que de esta manera se sientan cómodos con el nuevo sistema.

  Algunas de las preguntas a las que será imprescindible dar respuesta son las siguientes: ¿cómo afectará a su trabajo?, ¿qué cambios supondrá?, ¿el porqué de los mismos?, ¿cómo funciona el sistema?, ¿cómo les afectara a los empleados?, ¿cómo comunicárselo a los clientes? Los trabajadores deberán tener claro que nunca podrán culpar al sistema de la decisión tomada, puesto que éste se encuentra respaldado por la política de empresa, enseñándoles también a tratar con clientes insatisfechos

- La explicación a los clientes del funcionamiento del sistema. Debido al miedo que puede causar la implantación de nuevas tecnologías, es importante formar a los trabajadores para que sepan transmitir correctamente la información a los clientes, mediante folletos, páginas web y demás medios informativos, de esta manera se les explicara cual será el funcionamiento, los beneficios esperados para ellos mismos y así poder facilitar su entendimiento.

Una vez que ya se ha implantado el sistema habrá que probarlo antes de empezar a usarlo. En primer lugar, comprobar que funciona según lo diseñado y que los resultados obtenidos se asemejan a los esperados. Si apareciera algún error, se debería de corregir y comprobar de nuevo que todo funciona correctamente.

2.5. SOFTWARE

Otro punto a tener en cuenta son los equipos que se utilizarán, centrándonos en este caso en el software. Éste comprende el desarrollo del *scorecard* (panel de control) y el motor de decisión, que permitirá a la entidad llegar a una conclusión en cada caso. A continuación desarrollaremos estos dos aspectos.

En la actualidad existen dos tipos de software que permiten trabajar con el scorecard. El primero de ellos es más generalista, como son el SAS y el SPSS, y el segundo, más especializado en el *Credit Scoring*, como el SAS *Credit Scoring* Experian´s Sigma™, Scorex´s Toolbox™, existiendo otros muchos pese a que estos son las más frecuentemente utilizados. Hay que señalar también el software de código abierto R y la librería Credit Scoring de Gayler (2008).

A la hora de tomar una decisión sobre cuál es el programa que mejor se adapta a nuestra empresa, existirán una serie de factores que se deberán considerar como son: el coste del mismo, la facilidad para los empleados de su uso, servicio técnico ofrecido a la empresa, fácil adaptación a situaciones especiales y fácil comprensión.

Una vez elegido el programa a utilizar y definida la estrategia que va a seguir la entidad, llega el momento de ponerlo en marcha para poder tomar una decisión. Antiguamente los modelos de decisión se basaban en un análisis previo de las variables de cada individuo, generalmente llevado a cabo por el analista financiero o de riesgos pertinente. Éste emitía una valoración del individuo basándose en sus características y las del crédito solicitado, de esta manera la entidad podía decidir si aceptaba conceder el crédito o no. Este modelo de decisión cuenta con el principal inconveniente de la subjetividad y puede dar lugar a errores en la valoración.
En la actualidad, sin embargo, será el sistema de *scoring* implantado el que realice ese análisis de variables y datos dando una puntuación en cada caso sobre el solicitante. En función de la política de riesgos de la entidad o la situación en la que se encuentre, se estandarizará una puntuación que actuará como límite para la concesión o no del crédito. De esta manera se consigue eliminar la subjetividad de la que se hablaba previamente y se reduce el tiempo del proceso.

3. LA PRÁCTICA DEL *CREDIT SCORING*

Para la realización de esta parte del trabajo se ha mantenido contacto directo con dos instituciones financieras como son el Banco Santander y Liberbank, las cuales han servido de apoyo para la comprensión de esta práctica. Debido a la estandarización de los sistemas de *scoring* las dos entidades manejan programas similares, por lo que se describirá únicamente la parte correspondiente al Banco Santander de donde se obtuvo la mayor parte de la información.

3.1. BANCO SANTANDER

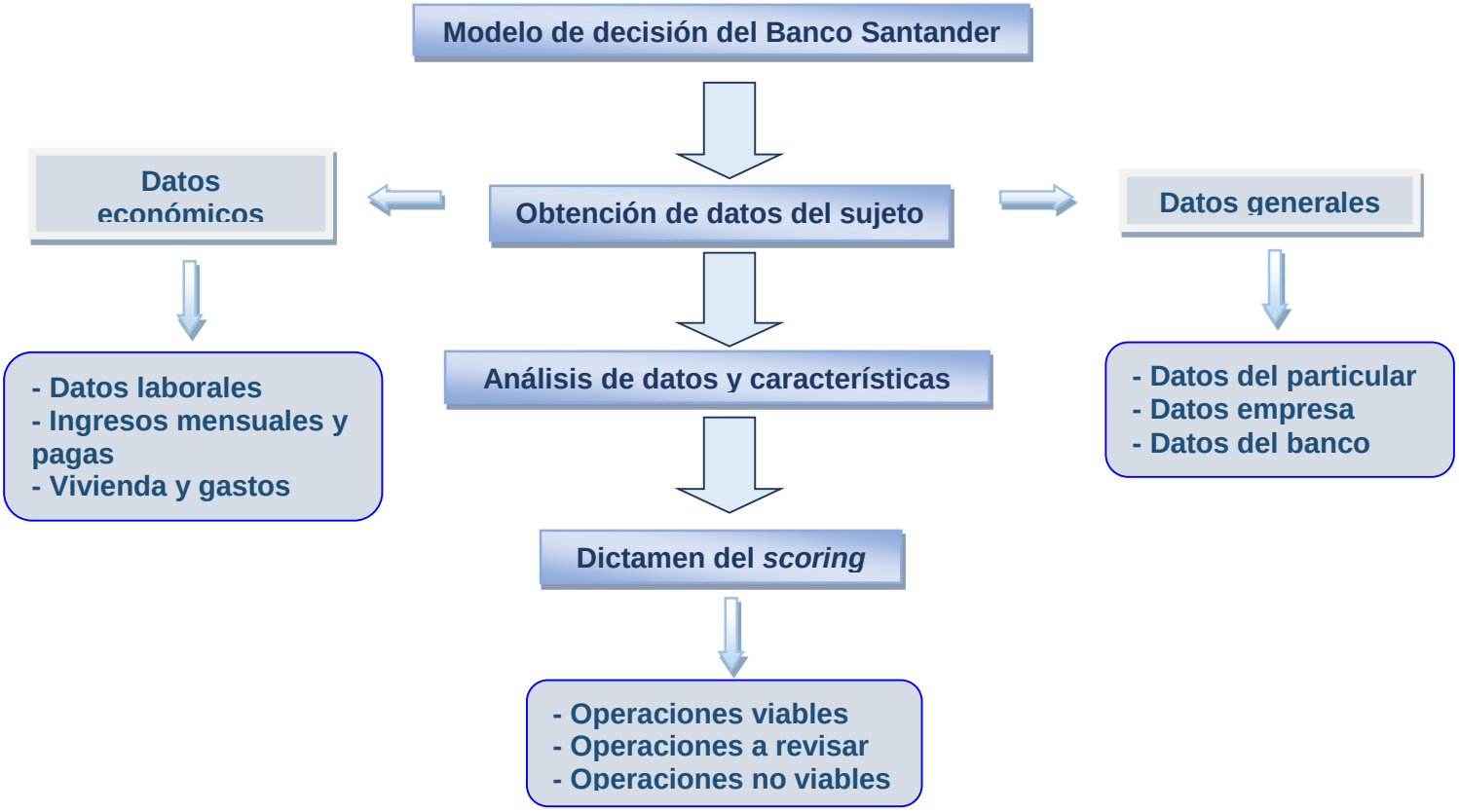

Figura 3.1: Modelo de *scoring* del Banco Santander

Tal como indica el manual que se nos cedió gentilmente, “Gestión de la morosidad y criterios de RR.EE en Banca comercial” (Recursos Humanos, Banco Santander 2009), la entidad financiera procederá a la obtención de los datos del sujeto para realizar el análisis y de esta manera, el *scoring* podrá determinar si el solicitante del crédito es apto o no apto. El sistema utilizado por el Banco Santander para esta labor de análisis de datos es el conocido como SAPA (Sistema de análisis de particulares) a través del cual podrá elaborar una opinión responsable acerca de los solicitantes, que será utilizada a la hora de tomar determinadas decisiones financieras o de riesgo, como es la concesión de un préstamo o crédito.
En primer lugar se imputarán los datos generales del cliente, que se dividirán en tres apartados: datos del particular, datos de la empresa y datos del banco.

- *Datos del particular*

  - Sexo
  - Estado civil
  - Nivel de estudios
  - Nacionalidad
  - Número de años viviendo en población actual
  - Número de hijos a su cargo
  - Número de años residiendo en su domicilio actual

- *Datos empresa*

  - Tipo de relación laboral (fijo, temporal, temporero ,autónomo, jubilado/pensionista)
  - Dirección
  - Nombre / Razón social
  - Teléfono
  - Otros (Complementar con los datos de la empresa en la que trabaja el sujeto al que le estamos realizando el SAPA)

- *Datos del banco*

  - Domiciliaciones (Nómina, recibos...)
  - Saldo medio otras cuentas
  - Antigüedad
  - Tarjetas de crédito

En segundo lugar, se imputarán los datos económicos del individuo, estos a su vez se agruparán en datos laborales, ingresos mensuales, vivienda habitual, otras propiedades, gastos mensuales y pagas extras.

- *Datos laborales*

  - Categoría profesional (Dato nómina, grupo de cotización
  - Actividad de la empresa (Objeto social empresa)
  - Años de antigüedad (antigüedad en la empresa actual)
  - Años antigüedad laboral (desde la primera vez que el trabajador fue dado de alta)

- *Ingresos mensuales*

  - Nómina neta (Líquido mensual a percibir)
  - Otros ingresos (Dietas, bonus o comisiones, arrendamientos…)
  - Procedencia

- *Pagas extras*

  - Número
  - Importe de cada una

- *Vivienda habitual*

  - Tipo de propiedad (propiedad sin cargos, etc.)
  - Valor
  - Hipoteca pendiente

- *Gastos mensuales*

  - Importe de alquiler
  - Importe de hipoteca vivienda habitual
  - Otros créditos/prestamos
  - Otros gastos especiales

Una vez que los datos económicos y personales han sido estudiados, se analizarán los riesgos de la operación con el dictamen del *scoring*. El *scoring* analiza la propuesta que se ha llevado a cabo y determinará la probabilidad de impago futura. Las políticas de la entidad financiera, en este caso el Banco Santander, determinará un nivel de impago máximo aceptable. Este nivel no es constante y dependerá del entorno económico, los resultados de la entidad y de otras variables que el Banco considere. En base a este nivel, el *scoring* devolverá el dictamen preciso sobre la operación y ésta podrá ser clasificada en: operaciones no viables, operaciones viables y operaciones a revisar

- *Operaciones viables*

  Las operaciones consideradas viables, presentan una baja probabilidad de impago por parte del cliente, por lo tanto, se considerará que el perfil de la operación está dentro de las políticas de riesgos definidas por el Banco.
  En este caso, la oficina que esté tramitando esta operación deberá autorizar la propuesta y facilitar su formalización, siempre y cuando la oficina no tenga información relevante contraria a la decisión del *scoring*. En cualquier caso, la decisión final siempre estará en manos del director de la oficina.

- *Operaciones no viables*

  Las operaciones consideradas no viables son aquellas que presentan una alta probabilidad de impago por parte del cliente, o bien, no están permitidas por las políticas de admisión del Banco. Estos supuestos se pueden dar, por ejemplo, cuando las hipotecas sean superiores a 40 años, el plazo de la operación mas la edad de los titulares sobrepase los 85 años, los datos de solicitud sean insuficientes o estén incompletos, etc. La oficina que está tramitando esta operación, deberá comunicar al cliente que la operación de crédito solicitada le ha sido denegada.
  Como ya se comentado previamente, el director de la oficina tiene la potestad suficiente para tomar una decisión acerca de la concesión del crédito, y podrá solicitar una reconsideración de la operación, siendo éste un procedimiento excepcional. Esta solicitud de reconsideración se puede deber a que el director disponga de una información adicional que no hay sido reflejada anteriormente y que pueda suponer una disminución del riesgo total de la operación.

- *Operaciones a revisar*

  Las operaciones consideradas a revisar presentan una probabilidad de incumplimiento no determinante para decidir la concesión del crédito. La solicitud ha sido valorada como admisible por el *scoring* pero incumple alguno de los criterios establecidos por la política del Banco. Estos casos se pueden dar cuando la capacidad de pago sea superior al 60% de la renta mensual neta destinada al pago de cuotas, a determinados informes negativos graves (tanto externos como internos) o cualquier otra circunstancia que la entidad precise.
  La oficina esta vez deberá delegar la decisión de la operación a un analista. Si el *scoring* que se ha obtenido es medio, se deberá reportar toda la información que se considere relevante al analista, para que éste realice otro análisis exhaustivo de la propuesta. En el caso de que el *scoring* obtenido sea alto, el analista únicamente deberá localizar el criterio de riesgo que haya sido incumplido.

4. APLICACIÓN PRÁCTICA

En esta sección se elabora un modelo real de *credit scoring* para poner en valor lo aprendido y desarrollado en los apartados anteriores. Inicialmente se intentó conseguir una muestra de datos en las dos oficinas bancarias consultadas, pero esta iniciativa se desestimó porque la información bancaria es claramente confidencial y estratégica. Por tanto, ante la imposibilidad de publicar los datos facilitados y de permitir a los lectores interesados la reproducción de los análisis realizados, se decidió recurrir a una base de datos pública y ampliamente analizada en la teoría de aprendizaje de máquina: la base de datos del crédito alemán. Hay que destacar también que dado el propósito del TFG, no importa tanto la muestra analizada como la metodología empleada en el análisis.

El conjunto de datos *German credit* está disponible en el Center for Machine Learning and Intelligent Systems de la Universidad de California en Irvine (Lichman 2013) y fue donada por el profesor Hofmann de la Universidad de Hamburgo en 1994. También está disponible en el Departamento de Estadística de la Universidad de Munich, así como en varios textos avanzados de estadística, entre los que cabe citar Ledolter (2013) y Kuhn y Johnson (2013). Una versión reducida se analiza en Farhmeir y Tutz (2001) y ha sido empaquetada en una librería de Farhmeir.R, por Halvorsen (2015).

La muestra de sección cruzada consta de 1000 casos (o solicitudes de crédito) y 20 atributos (o factores determinantes), y fue extraída usando un muestreo estratificado: 300 solicitudes extraídas del estrato de créditos malos y 700 del estrato de créditos buenos. La proporción que los créditos malos representan sobre los buenos es, por tanto, fija o no aleatoria.

Todo el análisis de datos que se presenta a continuación se ha realizado con el software gratuito Empiricus disponible en www.empiricus.unican.es.

4.1. ANÁLISIS EXPLORATORIO

En primer lugar, se realiza una descripción de estos atributos para familiarizarnos con sus valores, encontrar posibles errores y obtener una primera idea de la relación que guarda con la variable dependiente. Un análisis similar, aunque menos exhaustivo, puede encontrarse en la página web de Farhmeir y Tutz (2001), http://www.statistik.lmu.de/service/datenarchiv/kredit/kreditvar_e.html.

1º) Saldo en cuenta corriente. Se trata de una variable cualitativa ordinal que parece influir positivamente en la probabilidad de conceder un crédito. La tabla adjunta muestra las cuatro categorías de esta variable junto con los porcentajes de casos en cada estrato (clientes buenos o clientes malos). Además, se muestra una columna con las posibilidades de conceder un crédito en cada categoría. Estas posibilidades se calculan como la ratio de los porcentajes en los dos estratos. Por ejemplo, 0,44 = 19,86/45. También se presenta la ratio de posibilidades respecto a una categoría de referencia, en este caso, la categoría sin cuenta. Vemos que las posibilidades de obtener un crédito aumentan con el valor de la categoría. Las posibilidades de conseguir un crédito para los clientes con más de 200 marcos se multiplican por 7 cuando se compara con los clientes sin cuenta. Esta forma de interpretar la tabla de contingencia se aplica también para el resto de variables cualitativas.

Saldo	Categoría	Malo	Bueno	Odds	Ratio odds
Sin cuenta	1	45,00	19,86	0,44	
Sin saldo	2	35,00	23,43	0,67	1,52
< 200	3	4,67	7,00	1,50	3,40
>= 200	4	15,33	49,71	3,24	7,35

Tabla 4.1.1: Saldo en cuenta corriente

2º) Plazo del crédito. La duración del crédito se trata de una variable cuantitativa. Si comparamos la media de cada estrato podemos observar que es mayor en el correspondiente a clientes malos y también vemos que los dos extremos (Min y Max) son mayores en este estrato. De aquí se deduce que la variable influye negativamente en la probabilidad de conceder un crédito: cuanto mayor sea el plazo, menor probabilidad de que se conceda el crédito.

Plazo	Obs	Mean	Std	Min	Máx
Malo	300	24,86	13,28	6	72
Bueno	700	19,21	11,08	4	60

Tabla 4.1.2: Plazo del crédito

3º) Riesgo moral. La relación prestatario-prestamista es un ejemplo claro de información asimétrica donde las consecuencias del riesgo que asume el primero las sufre el segundo. Este riesgo moral se aproxima por una variable cualitativa ordinal que indica la historia crediticia del cliente. Las posibilidades de obtener un crédito aumentan con el valor de la categoría, las posibilidades de los individuos que han devuelto todos sus créditos es 8 veces más que los clientes que no tienen una historia clara de sus créditos.

Riesgo moral	Categoría	Malo	Bueno	Odds	Ratio odds
Pagos indecisos	0	8,33	2,14	0,26	
Cuenta problemática	1	9,33	3,00	0,32	1,25
Sin créditos previos	2	56,33	51,57	0,92	3,56
Sin problemas en créditos	3	9,33	8,57	0,92	3,57
Todos créditos devueltos	4	16,67	34,71	2,08	8,10

Tabla 4.1.3: Riesgo moral del individuo

4º) Uso del crédito. Se trata de una variable cualitativa nominal porque el orden de las categorías es arbitrario. Farhmeir y Tutz (2001) reducen esta variable policotómica a binaria con las categorías uso privado y uso profesional. En este caso, se observa que las posibilidades de conceder un crédito son superiores cuando se destina a usos privados.

Uso	Categoría	Malo	Bueno	Odds	Ratio odds
Otros	0	29,67	20,71	0,70	
Coche nuevo	1	5,67	12,29	2,17	3,11
Coche usado	2	19,33	17,57	0,91	1,30
Muebles	3	20,67	31,14	1,51	2,16
Radio / Televisión	4	1,33	1,14	0,86	1,23
Electrodomésticos	5	2,67	2,00	0,75	1,07
Reparaciones	6	7,33	4,00	0,55	0,78
Vacaciones	7	0,33	1,14	3,43	4,92
Formación	8	11,33	9,00	0,79	1,14
Negocios	9	1,67	1,00	0,60	0,86

Tabla 4.1.4: Uso del crédito

Uso	Categoría	Malo	Bueno	Odds	Ratio odds
Privado	1	57,33	69,29	1,21	
Profesional	2	42,67	30,71	0,72	0,60

Tabla 4.1.5: Codificación de Fahrmeir

5º) Cantidad del crédito. Comparando las medias de esta variable continua en cada estrato, podemos observar que es mayor en el correspondiente a los clientes malos. También presenta un máximo mayor este estrato. Se deduce, por tanto, que la variable influye negativamente en la probabilidad de conceder un crédito. Cuanta más cantidad se solicite, menor será la probabilidad de que se conceda.

Estrato	Obs	Mean	Std	Min	Máx
Malo	300	3938,13	3535,82	250	18424
Bueno	700	2985,44	2401,50	250	15857

Tabla 4.1.6: Cantidad del crédito

6º) Ahorros del individuo. Cantidad de ahorros que posee el solicitante del crédito. Se trata de una variable continua categorizada. Llama la atención que las posibilidades de conceder un crédito en la categoría 4 son mayores que en la categoría 5. Esto puede deberse a que el número de casos en la categoría 4 es bastante menor. Por esta razón se decide reducir esta variable policotómica a binaria con 1 si ahorros > 100 y 0 en otro caso.

Ahorros	Categoría	Malo	Bueno	Odds	Ratio odds
Sin ahorros	1	72,33	55,14	0,76	
100 <	2	11,33	9,86	0,87	1,14
100 < x < 500	3	3,67	7,43	2,03	2,66
500 < x < 1000	4	2,00	6,00	3,00	3,94
>= 1000	5	10,67	21,57	2,02	2,65

Tabla 4.1.7: Ahorros del individuo

7º) Empleo. Se trata de una variable cualitativa ordinal. De nuevo, llama la atención que las posibilidades de conceder un crédito en la categoría 4 son mayores que en la categoría 5. Esto puede deberse a que el número de casos en la categoría 4 es inferior. Por esta razón se convierte esta variable en binaria con 1 para más de un año y 0 en otro caso.

Empleo	Categoría	Malo	Bueno	Odds	Ratio odds
Sin empleo	1	7,67	5,57	0,73	
< 1 año	2	23,33	14,57	0,62	0,86
1 <= x < 4 años	3	34,67	33,57	0,97	1,33
4 <= x < 7 años	4	13,00	19,29	1,48	2,04
>= 7 años	5	21,33	27,00	1,27	1,74

Tabla 4.1.8: Empleo

8º) Cuota mensual. El porcentaje que la mensualidad (intereses más capital amortizado) representa en la renta del prestatario. Se trata de una variable cualitativa ordinal que parece estar mal codificada: ¡cuanto más alta es la cuota tanto más alta es la probabilidad de conceder un crédito! Se supone que se trata de un error de codificación, por lo que se considera que las categorías deberían ser: 1 para <20, 2 para (20, 25), …

Cuota	Categoría	Malo	Bueno	Odds	Ratio odds
35 <=	1	11,33	14,57	1,29	
25 <=< 35	2	20,67	24,14	1,17	0,91
20 <= < 25	3	15,00	16,00	1,07	0,83
< 20	4	53,00	45,29	0,85	0,66

Tabla 4.1.9: Cuota en porcentaje del ingreso disponible

9º) Aval para el crédito. Personas que avalan o solicitan con el individuo el crédito. Se trata de una variable cualitativa ordinal que parece influir positivamente en la probabilidad de conceder el crédito. Si se tiene un avalista la probabilidad de que se conceda un crédito es casi el doble frente a que nadie avale el crédito.

Aval	Categoría	Malo	Bueno	Odds	Ratio odds
Nadie	1	90,67	90,71	1,00	
Co-solicitante	2	6,00	3,29	0,55	0,55
Avalista	3	3,33	6,00	1,80	1,80

Tabla 4.1.10: Aval para el crédito

10º) Antigüedad de vivienda. Tiempo que el solicitante lleva viviendo en su residencia. Se trata de una variable cualitativa ordinal. Llama la atención que las posibilidades de conceder un crédito en la categoría 1 son superiores a las del resto de categorías. Por esta razón, se decide descartar esta variable.

Antigüedad vivienda	Categoría	Malo	Bueno	Odds	Ratio odds
< 1 año	1	12,00	13,43	1,12	
1 < x < 4 años	2	32,33	30,14	0,93	0,83
4 < x < 7 años	3	14,33	15,14	1,06	0,94
>= 7 años	4	41,33	41,29	1,00	0,89

Tabla 4.1.11: Antigüedad de vivienda

11º) Activos. Otros activos de los que disponga el solicitante. Puede considerarse como una variable cualitativa nominal puesto que el orden de las categorías parece arbitrario. Se puede observar que la categoría cuatro, propietario de casa o terreno, presenta una probabilidad bastante inferior que las otras variables, lo cual no tiene sentido, por lo tanto se puede intuir que esta variable no es relevante y será descartada.

Activos	Categoría	Malo	Bueno	Odds	Ratio odds
No activos	1	20,00	31,71	1,59	
Coche u otros	2	23,67	23,00	0,97	0,61
Seguros / planes de ahorro	3	34,00	32,86	0,97	0,61
Propietario de casa o terreno	4	22,33	12,43	0,56	0,35

Tabla 4.1.12: Activos

12º) Edad. Comparando las medias de edad en cada estrato, podemos observar que es algo mayor en el correspondiente a los clientes buenos. Parece razonable pensar que la probabilidad de conceder un crédito aumentará con la edad en un determinado tramo, pero en otro disminuirá. Por ejemplo, es más probable que se conceda un crédito a una persona de 40 años que a una de 20. En cambio, es más probable que se conceda un crédito a una de 70 que a una de 90. Esta situación suele modelizarse incluyendo el cuadrado de la edad como variable explicativa.

Edad (estrato)	Obs	Mean	Std	Min	Máx
Malo	300	33,96	11,23	19	74
Bueno	700	36,22	11,35	19	75

Tabla 4.1.13: Edad

13º) Otros créditos. El solicitante tiene que hacer frente a otros créditos vigentes. Podría considerarse como una variable cualitativa ordinal que parece influir positivamente en la probabilidad de conceder el crédito: cuanto menor sea la cuantía de los créditos vigentes, tanto mayor es la probabilidad de conceder un crédito nuevo.

Otros créditos	Categoría	Malo	Bueno	Odds	Ratio odds
En otros bancos	1	19,00	11,71	0,62	
En tiendas o centros comerciales	2	6,33	4,00	0,63	1,02
No tiene otros créditos	3	74,67	84,29	1,13	1,83

Tabla 4.1.14: Otros créditos

14º) Tipo de vivienda. Régimen de vivienda que tiene el solicitante. Las ratios de posibilidades toman valores contrarios a lo esperado: hay menos posibilidades de conceder un crédito a los propietarios de vivienda. Por estos motivos se descartará esta variable.

Tipo vivienda	Categoría	Malo	Bueno	Odds	Ratio odds
Apartamento social	1	23,33	15,57	0,67	
Apartamento alquilado	2	62,00	75,43	1,22	1,82
Propietario	3	14,67	9,00	0,61	0,92

Tabla 4.1.15: Tipo de vivienda

15º) Créditos previos misma entidad. Créditos previos que el solicitante haya tenido en la misma entidad incluyendo el actual. Se trata de una variable cualitativa ordinal, pero el orden de las categorías parece no influir en la probabilidad de conceder un crédito.

Créditos previos con mismo banco	Categoría	Malo	Bueno	Odds	Ratio odds
Uno	1	66,67	61,86	0,93	
Dos o tres	2	30,67	34,43	1,12	1,68
Cuatro o cinco	3	2,00	3,14	1,57	2,35
Seis o más	4	0,67	0,57	0,86	1,28

Tabla 4.1.16: Créditos previos misma identidad

16º) Ocupación. Situación laboral en la que se encuentra el solicitante. Se trata de una variable cualitativa nominal, puesto que el orden de algunas las categorías parece arbitrario. Como en otras variables, se observa que algunas ratios de posibilidades toman valores no esperados: las posibilidades de conceder un crédito son menores para una persona con empleo que para otra sin empleo. También se descarta este atributo.

Ocupación	Categoría	Malo	Bueno	Odds	Ratio odds
Sin empleo / sin residencia fija	1	2,33	2,14	0,92	
Sin empleo / con residencia fija	2	18,67	20,57	1,10	1,20
Con empleo	3	62,00	63,43	1,02	1,11
Ejecutivo / autónomo	4	17,00	13,86	0,82	0,89

Tabla 4.1.17: Ocupación

17º) Número de personas a su cargo. Las dos categorías de esta variable no parecen influir en la probabilidad de conceder un crédito.

Personas	Categoría	Malo	Bueno	Odds	Ratio odds
0 o 2	1	84,67	84,43	1,00	
3 o más	2	15,33	15,57	1,02	1,02

Tabla 4.1.18: Número de personas a mantener

18º) El solicitante dispone de teléfono o no. Las posibilidades de conceder un crédito son más altas cuando el cliente tiene teléfono.

Teléfono	Categoría	Malo	Bueno	Odds	Ratio odds
No	1	62,33	58,43	0,94	
Si	2	37,67	41,57	1,10	1,18

Tabla 4.1.19: El solicitante dispone de teléfono o no

19º) El individuo es extranjero. Las posibilidades de obtener un crédito son más altas para los extranjeros. Este resultado contrario a lo esperado parece deberse a la baja proporción de prestatarios extranjeros.

Extranjero	Categoría	Malo	Bueno	Odds	Ratio odds
No	1	98,67	95,29	0,97	
Si	2	1,33	4,71	3,54	3,66

Tabla 4.1.20: El individuo es extranjero o no

4.2. ESTIMACIÓN DEL MODELO

El análisis exploratorio de datos ha permitido realizar una selección preliminar de variables explicativas para estimar un modelo logit. Los resultados de la estimación por máxima verosimilitud se muestran en la tabla 4.2.1. Todas las variables destacadas en negro son significativas al nivel de significación del 5% (la ratio-z en valor absoluto es mayor que 1.96, el p-valor es menor que 5.0%, y el intervalo de confianza del 95% no incluye el valor 0); las variables destacadas en las variables en verde son claramente no significativas al nivel de significación del 15% o menor; las variables destacadas en rojo son claramente no significativas al nivel de significación del 15%.

Variable	Coeficiente	Error típico	z-ratio	p-valor (%)	[Intervalo	conf. 95,0%]
saldo==4	1,50	0,19	7,71	0,0	1,12	1,88
plazo	-0,03	0,01	-3,65	0,0	-0,05	-0,01
Ppag	1,02	0,27	3,73	0,0	0,49	1,56
uso==1	0,80	0,33	2,40	1,7	0,15	1,45
uso==0	-0,76	0,19	-3,93	0,0	-1,14	-0,38
uso==6	-0,88	0,36	-2,45	1,4	-1,59	-0,18
cantidad/100	-0,01	0,00	-2,87	0,4	-0,02	0,00
ahorros!=1	0,73	0,18	4,12	0,0	0,38	1,07
empleo>2	0,27	0,19	1,44	14,9	-0,10	0,64
cuota<3	0,69	0,19	3,71	0,0	0,33	1,06
Sexo	0,06	0,24	0,26	79,7	-0,41	0,54
Estc	-0,45	0,23	-1,93	5,3	-0,91	0,01
aval==3	0,93	0,40	2,34	1,9	0,15	1,71
edad	0,09	0,04	1,93	5,4	0,00	0,17
edad*edad	0,00	0,00	-1,61	10,7	0,00	0,00
otr*Creditos*==3	0,35	0,21	1,68	9,4	-0,06	0,75
prev*Credit*>1	0,05	0,17	0,27	78,7	-0,29	0,39
personas	-0,33	0,24	-1,36	17,3	-0,79	0,14
telef	0,27	0,18	1,52	12,8	-0,08	0,62
extranjero	1,33	0,61	2,19	2,9	0,14	2,52
constante	-3,43	1,10	3,12	0,2	1,27	5,59

Tabla 4.2.1: Estimación del modelo logit

A la vista de estos resultados, se llega a la conclusión de que las variables que la entidad financiera tendrá en cuenta a la hora de conceder el crédito a sus clientes, serán el saldo de la cuenta corriente, el plazo o duración del crédito, créditos previos pagados, el uso al que se destine, cantidad del crédito, ahorros del solicitante, porcentaje de la cuota disponible sobre los ingresos, estado civil, edad, si dispone de aval, si presenta otros créditos en el mismo banco y si es extranjero.

Es usual presentar un conjunto de estadísticos de diagnosis junto a la estimación del modelo, que se usan fundamentalmente para comparar modelos alternativos (tabla 4.2.2)

Log-Lik Intercept Only	-610.86	**Log-Lik Full Model**	-475.12
D(979)	950.25	**LR(20)**	271.48
		Prob > LR	0
McFadden's R2	0.222211	**McFadden's Adj R2**	0.187834
Maximum Likelihood R2	0.237751	**Cragg and Uhler's R2**	0.337101
McKelvey & Zavoina's R2	0.375953	**Efron's R2**	0.257888
Variance of y*	52.720	**Variance of error**	32.900
Count R2	0.765000	**Adj Count R2**	0.216667
AIC	0.992247	**AIC*n**	992.25
BIC	-5812.45	**BIC'**	-133.33

Tabla 4.2.2: Estadísticos de diagnosis

4.3. EVALUACIÓN DEL MODELO

Cuando se utiliza un modelo de *credit scoring* para decidir si se acepta o se rechaza una petición crediticia, pueden cometerse dos tipos de error: (1) el "falso positivo", o error de conceder un crédito a un mal cliente, y (2) el "falso negativo", o error de denegar un crédito a un buen cliente. También hay dos tipos de acierto: el "verdadero positivo", o acierto al conceder un crédito a un buen cliente, (2) el "verdadero" negativo, o acierto al denegar un crédito a un mal cliente, "verdadero negativo". Estos errores y aciertos pueden presentarse en una tabla de contingencia que cruce los dos tipos de decisiones (rechazar o aceptar una petición crediticia) y los dos tipos de prestatarios (malos y buenos).

		Prestatarios	
Decisión		Malos	Buenos
	Rechazar	VN	FN
	Aceptar	FP	VP

Tabla 4.3.1: Tabla de contingencia teórica

En esta tabla, VN y FP son el número de peticiones crediticias de malos prestatarios que han sido acertadamente rechazadas y erróneamente aceptadas, respectivamente; VP y FN son es el número préstamos de buenos clientes que han sido acertadamente aceptadas y erróneamente rechazadas, respectivamente. De aquí, se definen la tasa de falsos positivos para prestatarios malos, TFP = FP/(VN+FP), y la tasa de falsos negativos para los prestatarios buenos, TFN = FN/(FN+VP). Las correspondientes tasas complementarias se denominan sensibilidad, 1-TFN, y especificidad, 1-TFP. Mientras que la sensibilidad en este marco informa de la capacidad del modelo para reconocer clientes buenos, la especificidad indica la capacidad para identificar clientes malos. Por tanto, y como señalan Garanin y otros (2014), un modelo con alta especificidad corresponde a una política crediticia conservadora por parte del banco, mientras que un modelo con alta sensibilidad se corresponde con una política crediticia conservadora.

La sensibilidad y especificidad vienen determinadas por el punto de corte con el que se compara la probabilidad estimada de devolver el crédito. Así, si Prob(Yi=1) >c, se concede el crédito; en caso contrario, se deniega.

Cuando no se dispone de información cuantitativa sobre el coste de los dos errores, el punto de corte se suele fijar en 0,5. Si se usa este valor, la tabla de contingencia del modelo logit para los datos *German credit* es la siguiente:

	Malos	Buenos	Predicciones	Aciertos (%)
Rechazar	140	75	215	0,651163
Aceptar	160	625	785	0,796178
Observaciones	300	700	1000	
Aciertos (%)	0,466667	0,892857		

Tabla 4.3.2: Tabla de contingencia del modelo logit, punto de corte=0,5

Se observa que la sensibilidad, 0,89, es mayor que la especificidad, 0.46. Ahora bien, si se fija un punto de corte más conservador, mayor que 0,5, estas tasas cambiarán. Por ejemplo, en la página web de la UCI se dice que el coste de error FN es 5 veces el del FP. De aquí, si q es la probabilidad de que el cliente no devuelva el crédito, el coste esperado del FN es 5q, mientras que el coste esperado del FP es (1-q). Así, se cumplirá

que el primer coste es menor que el segundo, 5q < (1-q), cuando q < 1/6, lo que conduce a un punto de corte c = 5/6. La correspondiente tabla de aciertos y fallos es ahora

	Malos	Buenos	Predicciones	Aciertos (%)
Rechazar	269	335	604	0,651163
Aceptar	31	365	396	0,796178
Observaciones	300	700	1000	
Aciertos (%)	0.896667	0.521429		

Tabla 4.3.3: Tabla de contingencia del modelo logit, punto de corte= 5/6

Se observa que la sensibilidad, 0,52, es mayor que la especificidad, 0.90. Como distintos valores del punto de corte conducen a diferentes tasas de sensibilidad y especificidad, se recomienda construir la curva ROC (acrónimo de la denominación inglesa *Receiver Operating Characteristics*). Esta curva representa la sensibilidad, 1-TFN, frente a 1-especificidad o TFN, para distintos puntos de corte. Se usa para evaluar la capacidad predictiva del modelo. El modelo ideal corresponde al punto superior izquierda en el que la sensibilidad = 1 – especificidad = 1, es decir, el que clasifica correctamente a todos los clientes. Por el contrario, el punto inferior izquierdo corresponde al peor modelo posible, el que no clasifica correctamente a ningún cliente (TVP=TVN=0). Tampoco son buenos modelos los que conducen a un punto sobre la línea de 45º, en el que las dos tasas de fallos suman la unidad, TFP+TFN = 1. En el punto extremo, (1,1), todos los casos se clasifican como buenos. De aquí, cuánto más próxima se encuentre la curva ROC del punto (0,1) mejor será el modelo.

El gráfico 4.3.4 muestra la curva ROC para el modelo logit de credit scoring. El área bajo la curva, calculada con la regla del trapecio, es 0.86, que puede considerarse como una adecuada capacidad predictiva.

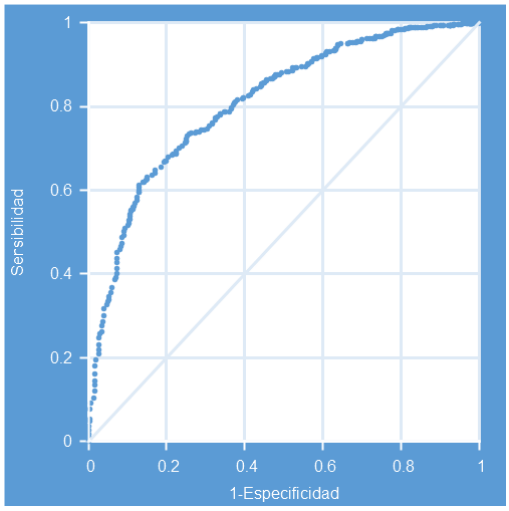

Gráfico 4.3.3: curva ROC para el modelo logit

4.4. USO DEL MODELO

Una vez estimado el modelo, puede usarse para estimar la probabilidad de que un determinado cliente devuelva el crédito. Por ejemplo, la tabla x muestra las características de dos clientes que solicitan un crédito. Presentan las mismas características, salvo el plazo y la cuantía. El indicador para el cliente A es 3.12; para el B, 1.17. Estos indicadores se convierte en probabilidades utilizando la función de distribución logística: 1/(1+exp (-3,12)) = 0.94 y 1/(1+exp(-1.17)) = 0.76. Como se

esperaba la probabilidad de conceder un crédito disminuye con la cantidad y con la cuota.

Variable	Coeficiente	Cliente A	Cliente B
Saldo==4	1,4985	1	1
plazo	-0,03073	21	63
Ppag	1,0284	1	1
uso==0	-0,784168	0	0
uso==1	0,756193	1	1
uso==6	-0,913979	0	0
cantidad/100	-0,011079	30	90
ahorros!=1	0,720216	1	1
ante>2	0,273558	1	1
cuota<3	0,673041	1	1
Estc	-0,449074	1	1
aval==3	0,914547	1	1
edad	0,075711	30	30
edad*edad	-0,000726	900	900
ocreditos==3	0,351722	1	1
telef	0,281644	1	1
extranjero	1,3123	0	0
1	-3,5612	1	1
		3,127777	1,172377

Tabla 4.3.4: Tabla de probabilidad de los clientes

5. CONCLUSIONES

El tema que se ha desarrollado es enormemente extenso, puesto que son muchos los trabajos de investigación realizados con diferentes aportaciones en cada uno de ellos. Se pueden extraer varias conclusiones acerca de lo explicado en el presente documento:

- Este sistema facilita el uso eficiente de la gestión de riesgos en las entidades bancarias, puesto que permite mejorar la calidad de los créditos.

- Permite otros usos de sus resultados, puesto que la clasificación de los clientes como aptos o no aptos es útil para campañas comerciales dentro de la propia entidad, scoring proactivo.

- Pese a ser un sistema con un elevado coste, la aplicación del mismo conlleva una significativa reducción de los tiempos de actuación y decisión a la hora de estimar y evaluar el riesgo del crédito.

- Los resultados obtenidos por este modelo facilitan a la entidad el trabajo, ya que se elimina el factor de subjetividad en las valoraciones.

- Este sistema tiene una gran dependencia de los datos, puesto que para emitir una valoración lo más correcta, necesita disponer de la mayor información posible acerca del cliente.

- No todas las variables tiene el mismo nivel de significación por lo que será necesario realizar un análisis exhaustivo de éstas, con la finalidad de encontrar posibles errores y obtener una primera idea de la relación que guarda con la variable dependiente.

Con este trabajo queremos seguir dejando la puerta abierta a las futuras líneas de investigación que puedan sobrevenir al respecto tanto por tratarse de un tema amplio, como por la luz que puede arrojar sobre el control de riesgos en las entidades financieras.

6. BIBLIOGRAFÍA

ABDOU, H.A.; POINTON, J. 2011. Credit Scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management,* **18** (2-3), pp. 59-88. [Consulta: 5 mayo 2015]. Disponible en: http://onlinelibrary.wiley.com/doi/10.1002/isaf.325/abstract.

ANDERSON, R. 2007. *The Credit Scoring Toolkit*: *Theory and Practice for Retail Credit Risk Management and Decision Automation*. New York: Oxford University Press. ISBN 978-0-19-922640-5.

BARTOLOZZI, E [et al.]. 2008. Credit Scoring Modelling for Retail Banking Sector. En: Universidad Complutense de Madrid (ed.). *II Modelling Week*, June 16th-24th, Madrid, pp. 1-9. [Consulta: 19 mayo 2015]. Disponible en: http://www.mat.ucm.es/momat/2008mw/creditscoring.pdf

CHYE KOH, H.; CHIN TAN, W.; PENG GOH, C. 2006. A two-step method to construct *Credit Scoring* models with data mining techniques. *International Journal of Business and Information*, **1** (1), pp. 96-118.

COSTA COR, T.; BOJ DEL VAL, E.; FORTIANA GREGORI, J. 2012. Bondad de ajuste y elección del punto de corte en regresión logística basada en distancias. Aplicación al problema de *Credit Scoring. Anales del Instituto de Actuarios Españoles,* **18**, pp. 19-40.

FARHMEIR, L; TUTZ, G. 2001. Multivariate Statistical Modelling Based on Generalized Linear Models. 2ª ed. New York: Springer. ISBN 978-1-4419-2900-6

FINLAY, S. 2010. *Credit Scoring, Response Modelling and Insurance rating: A Practical Guide to Forecasting Consumer Behaviour.* Great Britain: Palgrave Macmillan. ISBN 978-0-230-57704-6.

GARANIN, D.A.; LUKASHEVICH, N.S.; SALKUTSAN, S.V. 2014. The evaluation of Credit Scoring models parameters using ROC curve analysis. *World Applied Sciences Journal*, **30** (8), pp. 938-942. ISSN 1818-4952. [Consulta: 15 septiembre 2015]. Disponible en: http://www.idosi.org/wasj/wasj30(8)14/2.pdf

GAYLER, R. 2008. *Credit Risk Analystics Occasional Newsletter* . Great Britain.

HOFMANN (1994) LICHMAN, M. 2013. UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. [Consulta: 15 abril 2015]. Disponible en: https://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)

LEUNG, K [et al.]. 2008. Building a Scorecard in Practice. En: CHUNG-MING, O. (ed.). *Proceedings of the 7th International Conference on Computational Intelligence in Economics and Finance,* December 5th-7th, Taiwan, pp.1-7. [Consulta: 10 mayo 2015]. Disponible en: http://www.aiecon.org/conference/2008/CIEF/Building%20a%20Scorecard%20in%20Practice.pdf

THOMAS, L.C.; EDELMAN, D.B.; CROOK, J.N. 2002. *Credit Scoring and Its Applications.* Philadelphia: Society for Industrial and Applied Mathematics. ISBN 0-89871-483-4.

AGRADECIMIENTOS

No podría finalizar este trabajo sin dejar constancia de mi más profundo agradecimiento al director de este Trabajo Fin de Grado, el profesor José Luis Gallego Gómez. Gracias por su inestimable paciencia, por su actitud positiva, por su entusiasmo por la docencia y lograr transmitírnoslo a los alumnos generación tras generación. Ha sido un placer aprender de usted.

Gracias a mis padres y a mi hermana por no dejar de apoyarme nunca.

A Cristina que espero poder devolverle un día todo lo que me ha ayudado.

Patricia, Lucía, Leticia y Borja que han sido los mejores compañeros que podía haber tenido en esta carrera.