



**BENEMÉRITA UNIVERSIDAD  
AUTÓNOMA DE PUEBLA**

*FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS*

*Riesgo de Crédito: Análisis Mediante Redes  
Neuronales*

**T E S I S**

*que para obtener el título de:*

**LICENCIADO EN ACTUARÍA**

*presenta:*

*ADRIANA HERRERA MARTÍNEZ*

*Directores de tesis:*

*DRA. HORTENSIA REYES CERVANTES  
DR. FRANCISCO SOLANO TAJONAR SANABRIA*

*PUEBLA, PUE.*

*SEPTIEMBRE 2017*



***BENEMÉRITA UNIVERSIDAD  
AUTÓNOMA DE PUEBLA***

*FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS*

*Riesgo de Crédito: Análisis Mediante Redes  
Neuronales.*

*T E S I S*

*que para obtener el título de:*

***LICENCIADO EN ACTUARÍA***

*presenta:*

*ADRIANA HERRERA MARTÍNEZ*

*Directores de tesis:*

*DRA. HORTENSIA REYES CERVANTES  
DR. FRANCISCO SOLANO TAJONAR SANABRIA*

*PUEBLA, PUE.*

*SEPTIEMBRE 2017*



*Dedico esta tesis con todo mi cariño:*

A mis padres por enseñarme a  
siempre perseguir mis sueños.  
Y a mi hermano por inspirarme  
a ir más lejos.



# *Agradecimientos*

A mi madre por ayudarme incondicionalmente y ser mi roca, a mi padre por apoyarme siempre y por ser un ejemplo de que se puede perseverar en la adversidad, a mi gemeli por asimilarme y comprenderme, y a toda mi familia por su apoyo.

A mis directores de tesis. Dra. Hortensia Reyes Cervantes por su ayuda a lo largo de este proceso, por compartir sus conocimientos conmigo, por su paciencia, perseverancia y su apoyo durante mi carrera. Dr. Francisco Solano Tajonar Sanabria por su accesibilidad para ayudarme en este proceso, por sus consejos e inestimable apoyo.

A mis sinodales. Dr. Hugo Adán Cruz Suárez, Dra. Gladys Linares Fleytes y M. C. Brenda Zavala López, por sus comentarios y observaciones, por su disponibilidad y tiempo para la revisión de este trabajo.

A todos y cada uno de los profesores que durante mis años de estudio compartieron sus conocimientos y me ayudaron a desarrollarme intelectual y personalmente, y me inspiraron a mejorar y retarme para poder superarme.

A todos mis amigos que hice durante estos años, por esas largas horas de estudio sobretodo en teoría del juego, por todas las risas que compartimos, por el tiempo que invertimos juntos para mejorar, por las celebraciones que hicimos y sobretodo por ayudarme a mejorar como ser humano.

Y a todos las personas que han formado parte de mi vida.





# Índice

<b>1. Introducción</b>	<b>1</b>
<b>2. Preliminares</b>	<b>7</b>
2.1. Crédito . . . . .	7
2.2. Aspectos Generales del Credit Scoring . . . . .	8
2.2.1. Definición y Conceptos Básicos . . . . .	9
2.2.2. Ventajas del Credit Scoring y Limitaciones . . . . .	10
2.3. Variables Empleadas . . . . .	11
2.3.1. Construcción del Modelo . . . . .	15
2.4. Modelos de Credit Scoring . . . . .	16
2.4.1. Juicio Humano vs Modelos Scoring . . . . .	17
2.4.2. Análisis Discriminante . . . . .	20
2.4.3. Regresión Logística . . . . .	22
2.4.4. Árboles de Regresión y Clasificación (CART) . . . . .	28
2.4.5. Redes Neuronales . . . . .	33
<b>3. Problemática</b>	<b>41</b>
3.1. Crisis Financiera de Taiwan . . . . .	43
3.2. Descripción de la Base de Datos . . . . .	44
3.2.1. Descripción de las Variables . . . . .	45
<b>4. Una aplicación del Credit Scoring</b>	<b>49</b>
4.1. Resultados . . . . .	50

---

<b>5. Conclusiones</b>	<b>57</b>
<b>A. Estadística</b>	<b>61</b>
A.1. Distancia de Mahalanobis . . . . .	61
A.2. Distribución Normal Multivariada . . . . .	62
A.3. Pruebas Aplicadas en el Análisis Discriminante . . . . .	62
<b>B. Tablas de Frecuencias de la Base de Datos</b>	<b>65</b>
<b>C. Red Neuronal en SPSS</b>	<b>67</b>
<b>D. Glosario</b>	<b>77</b>

# Capítulo 1

## Introducción

A lo largo de los años el aumento en el uso de créditos ha llevado a la industria a desarrollar métodos para determinar a qué personas físicas o morales otórgales uno. Para una institución financiera es esencial poder reducir los riesgos que puedan producir pérdidas debido a una mala selección de clientes, ya que si los créditos otorgados no son pagados estos se traducen en pérdidas económicas para estas instituciones.

De la Encuesta Nacional de la Inclusión Financiera (ENIF) realizada por la Comisión Nacional Bancaria y de Valores (CNBV) y el Instituto Nacional de Estadística y Geografía (INEGI) en enero de 2016, en México se reportó un aumento en el número de adultos que cuentan con algún producto financiero. La CNBV registro que al cierre de febrero de 2016 circulaban en el mercado alrededor de 22 millones de tarjetas de crédito, 192 mil 722 más de las que se registraron en febrero de 2015, lo cual denota una recuperación ya que de 2014 a 2015 los bancos perdieron alrededor de medio millón de tarjetas de crédito. A su vez el presidente de la CNBV, Jaime González Aguadé observa que se ha avanzado mucho en la inclusión de cuentas bancarias a lo largo de los tres años que dura la encuesta y que a pesar de que más de la mitad de la población adulta no cuenta con ninguna, se ha observado que aún hay áreas de oportunidad

en la que se debe seguir trabajando [2].

Igualmente, información proporcionada por el Banco de México y la CNBV, muestra que en enero de este año se registro el nivel más alto en el monto de cartera vencida a préstamos al consumo, particularmente hubo un aumento del 8.61 % en la cartera vencida de usuarios de tarjetas de crédito respecto a enero de 2016, lo cual denota la importancia de el problema al que se enfrentan las instituciones financieras al momento de decidir a que personas otorgales un crédito [8].

Los modelos de Credit Scoring son herramientas estadísticas que se han desarrollado a lo largo del tiempo para poder realizar la clasificación de clientes en buenos o malos. Aunque en un inicio la decisión de otorgar un crédito o préstamo estaba bajo la decisión de un gerente de crédito (o la persona en la institución a cargo del área de préstamos), el cual, principalmente mediante su juicio, determinaba a quien se le otorgaba un préstamo. Hoy en día, es cada vez más común el uso de los modelos de Credit Scoring, para llevar a cabo estos procesos (de aprobación de créditos) de manera más objetiva y precisa.

Aunque estos modelos cuentan con un buen respaldo matemático no son capaces de determinar de manera precisa el comportamiento exacto que cada solicitante (a un crédito) tendrá. Sin embargo, otorga un buen panorama de lo que se espera sea el comportamiento de los solicitantes con que cuenta la institución.

Estos modelos se basan en la información con la que cuentan las instituciones financieras sobre los solicitantes a créditos. Estas características de interés varían de una institución a otra y del tipo de solicitante (persona física o moral), también, generalmente se incluyen datos demográficos como edad, género, etc., así como datos financieros de los solicitantes como ingresos, historial de crédito, morosidad en los pagos, etc.

Para determinar a los buenos clientes las instituciones financieras cuentan con una fórmula con parámetro desconocido, que estiman con base en la información que tienen de los solicitantes y la experiencia que poseen (de créditos anteriores); sin embargo, también es posible estimarse con información de otras instituciones financieras (información externa). A partir de este parámetro se podrá determinar las probabilidades de los solicitantes de cometer un default.

Con base en esta probabilidad las instituciones financieras determinan el puntaje (score) con el cual se le otorgará un crédito a un solicitante. Este puntaje mínimo que debe cumplir un solicitante para ser considerado como buen cliente es determinado por cada institución financiera con base en el nivel de riesgo que está dispuesta a afrontar.

De esta manera, al tener una fórmula directa solo es necesario que los prestamistas ingresen la información necesaria de los solicitantes y podrán determinar si el cliente es bueno o malo para formar parte de la cartera de préstamos. Este tipo de métodos es más eficiente al poder determinar de forma rápida la clasificación de los clientes.

Uno de los mayores problemas con los que se cuenta para el desarrollo de nuevos sistemas de Credit Scoring es la escasez de información pública disponible (la información disponible es muchas veces desactualizada y en algunos casos incompletas). Esto es debido a la confidencialidad que las instituciones financieras deben mantener con sus solicitantes y clientes.

El uso de estos sistemas no solo se limita al otorgamiento de tarjetas de crédito o de préstamos, sino que también se utiliza en distintos tipos de financiación tanto en hipotecas, como para los seguros, incluso para servicios como rentas de teléfonos; cada vez es más común su uso en distintas áreas [9].

La finalidad de este trabajo es aprender la metodología de Credit Scoring, y aplicarlo a una base de datos real. En este proceso se pudo conseguir la base de datos de un sitio de internet de la Universidad de California, otorgada por el Ph.D. I-Cheng Yeh. La base de datos contiene la información de 30,000 clientes de un banco de Taiwan, así como 23 variables explicativas con información socioeconómica de los clientes de Abril a Septiembre de 2005.

Para realizar la aplicación se uso el método de redes neuronales, una técnica representativa de la minería de datos, que tiene su origen en diversas disciplinas: neurociencias, matemáticas, estadística, física, ciencias computacionales e ingeniería. Las redes neuronales se aplican en diversos campos tales como modelación, análisis de series de tiempo, reconocimiento de patrones, procesamiento de señales, entre otros. Debido a su poder predictivo y su eficacia demostrada en el desarrollo de Credit Scoring, se uso la arquitectura Perceptrón Multicapa (MLP) la cual, además tiene la ventaja de ser un modelo desarrollado en diversos software.

Esta técnica a sido aplicada por diversas empresas, que han creado sus sistemas basados en redes neuronales tales como American Express (para detección de fraudes en tarjetas de crédito), Lloyds Bowmarker Motor Finance (para sistemas de Credit Scoring en decisiones de financiamiento automotriz), Security Pacific Bank (para sistemas de Credit Scoring para préstamos a empresas pequeñas), entre otras [23].

Esta tesis está dividida en 5 capítulos.

En el Capítulo 1, se da una introducción a la temática del Credit Score y su utilidad en las empresas de crédito. Metodología que ha ido tomando popularidad en otras áreas de trabajo.

En el Capítulo 2 se dan a conocer algunos conceptos de Credit Sco-

ring, definiciones de crédito, así como el tipo de variables usadas para la determinación de Credit Scoring. También se da una breve introducción de diversos métodos scoring.

Posteriormente en el Capítulo 3 se habla de la base de datos usada, también se dan a conocer los antecedentes históricos relevantes en la economía de Taiwán durante la fecha de recolección de la información.

La aplicación del método se explica en el Capítulo 4, en él se encuentra un resumen de los resultados que se obtuvieron durante la aplicación del método a la base de datos.

Finalmente, tenemos el Capítulo 5 con las conclusiones a las que se llegaron después de la aplicación del método.





# Capítulo 2

## Preliminares

### 2.1. Crédito

Etimológicamente la palabra crédito viene del latín *credere*, que significa confiar [26].

La Real Academia Española (RAE) define un crédito como:

“Cantidad de dinero u otro medio de pago que una persona o entidad, especialmente bancaria, presta a otro bajo determinadas condiciones de devolución”[20].

A esta definición se puede agregar, que estas condiciones generalmente son: que el crédito debe ser pagado con intereses y en pagos (generalmente hechos en tiempos periódicos).

Un préstamo consiste en la entrega física o moral de un bien o cierta cantidad de dinero que una persona (física o moral) le otorga a otra [26].

Los créditos se pueden clasificar en 3 tipos:

*Créditos de consumo o créditos comerciales.* Son los créditos que otorgan las empresas para el consumo de bienes o servicios de uso

personal, que serán pagados en plazos previamente determinados.

*Créditos Empresariales.* Se otorgan a empresas que requieren materia prima, insumos, servicios, etc. Se solicita el bien o servicio a otras empresas a crédito para poder continuar su actividad empresarial, realizando convenios para cubrir el adeudo en un futuro.

*Créditos Bancarios.* Estos son concedidos por instituciones financieras a personas físicas o morales que requieran recursos para financiar sus actividades, tales como la adquisición de bienes, servicios, pago de deudas, entre otros.

Las tarjetas de crédito son uno de los productos que las instituciones financieras poseen. Estas son, el medio de pago que un banco o un comercio otorga a las personas, para que puedan realizar (sin utilizar dinero en efectivo) el pago de bienes y servicios de forma inmediata que posteriormente se liquidará, a veces en pagos, con el correspondiente pago de intereses.

Existen tres tipos de intereses que utilizan las tarjetas de crédito:

- a) **Tasa revolvente.** Es la tasa que se cobra por el uso de la tarjeta y está dada en función a la TIIE (Tasa de Interés Interbancaria de Equilibrio).
- b) **Tasa fija.** No cambia durante el plazo para la liquidación de la deuda, depende del cliente.
- c) **Tasa cero.** No se cobra ninguna tasa de interés el cliente solo debe cubrir el monto del servicio adquirido en un periodo determinado [17].

## 2.2. Aspectos Generales del Credit Scoring

Hoy en día es más común el uso de Credit Scoring para el proceso de aprobación de un crédito. Este método fue introducido en los 50's, y

actualmente es ampliamente usado en los créditos al consumo, especialmente en el área de tarjetas de crédito, de igual forma es cada vez más usado para otorgar hipotecas y créditos a PyMEs (Pequeñas y medianas empresas) [15].

### **2.2.1. Definición y Conceptos Básicos**

En el artículo de Loretta J. Mester llamado “What's the point of Credit Scoring”, se definió al Credit Scoring como:

“... un método de evaluación del riesgo de crédito de solicitudes de préstamo. Usando información histórica y técnicas estadísticas, el Credit Scoring trata de identificar los efectos de varias características de los solicitantes en mora o en fallo. El método produce un “score” que un banco puede usar para clasificar sus solicitudes de crédito o deudores en términos de riesgo. Para construir un modelo Scoring, o “score-card”, los desarrolladores analizan datos históricos sobre el desempeño hecho en préstamos previos para determinar cuáles características de los deudores son útiles en la predicción si el desempeño será bueno. Un modelo bien diseñado, debe dar un alto porcentaje de scores altos a deudores cuyos préstamos se desempeñarán bien y un alto porcentaje de scores bajos a deudores que no se desempeñarán bien. Pero ningún modelo es perfecto, y algunas malas cuentas recibirán scores más altos que algunas buenas cuentas”[15].

De manera más concreta Hand y Henley definieron el Credit Scoring como:

“... el término usado para describir los métodos estadísticos formales usados para clasificar solicitudes de crédito en clases buenos o malos”[10].

Para construir un modelo Scoring la institución financiera debe determinar el score de corte (o cut-off), es decir, el score mínimo que debe tener un solicitante para ser clasificado como “bueno”. A mayor riesgo,

mayor ganancia; la institución financiera debe buscar un punto de equilibrio entre el riesgo que está dispuesto a aceptar y las ganancias que espera obtener y así determinar el score de corte.

### 2.2.2. Ventajas del Credit Scoring y Limitaciones

La principal ventaja que presenta el uso de los modelos de Credit Scoring es su objetividad y la precisión que se obtiene en la clasificación de los solicitantes de crédito. Una clasificación precisa, se considera beneficiosa para ambas partes: al prestamista le beneficia en términos de aumento en sus ganancias o en reducción de pérdidas, y al solicitante le beneficia al evitar sobre compromisos [10].

Otro de los atractivos que presenta el uso del Credit Scoring, es que reduce los tiempos del proceso de aprobación de crédito. Este tiempo ahorrado significa menos gastos para la institución financiera y también representa un beneficio para el solicitante (teniendo una pronta respuesta a su crédito).

Una de las principales limitaciones que se presenta en la clasificación de solicitantes, con independencia del enfoque que se ocupe, es la no aleatoriedad en las muestras. En su mayoría, los trabajos realizados en ésta área han utilizado muestras truncadas, es decir, dado que estas muestras se forman solo con los créditos que son aceptados, esto debido a la imposibilidad de obtener datos de créditos no concedidos.

Las limitaciones que presenta este tipo de métodos, es que la precisión de los sistemas de Credit Scoring para grupos no representados es todavía una pregunta abierta. La precisión es una característica muy importante del Scoring. Aunque el prestamista pueda reducir sus costos de evaluación de solicitudes de préstamos debido a su uso, si los modelos no son precisos, estos costos ahorrados podrían ser consumidos por malos desempeños en los préstamos. La precisión de un sistema Scoring

depende del cuidado con el que se desarrolla y se construye. Los datos en los cuales el sistema se basa necesitan ser una muestra rica en ambos tipos de desempeños los buenos y los malos.

También es importante saber que un modelo de Credit Scoring no le va a decir a una institución financiera con certeza cómo será el desempeño futuro de un individuo. En la clasificación de solicitudes se pueden dar dos clases de errores, clasificar solicitantes malos como buenos o clasificar a solicitantes buenos como malos, siendo el primer tipo de error el que representará más pérdidas. Ningún modelo Scoring puede prevenir este tipo de errores, pero un buen modelo debe ser capaz de predecir con exactitud el desempeño promedio hecho por grupos de individuos quienes comparten valores similares de los factores identificados como relevantes para la calidad crediticia [15].

## 2.3. Variables Empleadas

En la elaboración del Credit Scoring hay 3 enfoques comúnmente usados para la selección de variables:

- El primer criterio se basa en hacer uso del conocimiento especializado, la experiencia y la opinión, sobre los datos y las variables, lo cual es buen complemento en la manipulación de la estadística formal. Esto impedirá que variables no predictivas sean incluidas, por razones históricas, de esta manera se puede justificar la selección de variables. Ya que es necesario poder justificar el sistema a los usuarios.
- Otro criterio es el uso de un procedimiento de estadístico escalonado [10]. El método Forward, consiste en ir agregando variables secuencialmente al modelo, de modo que el poder predictivo del modelo mejore con cada variable agregada. El método Backward, comienza con un modelo complejo y se eliminan variables consecutivamente, de tal forma que con cada eliminación cause el menor

daño al modelo y lo mejore, el método se detiene cuando las eliminaciones den un ajuste con menor poder predictivo al anterior. Algunos softwares tienen opciones adicionales para la selección de variables [1].

- Por último, la selección de variables individuales usando una medida de diferencias entre la distribución de riesgos buenos y malos sobre esa variable. El valor informativo es una medida de uso común, definida como:

$$\sum_j (p_{ij} - q_{ij}) w_{ij}, \quad \text{con } i = 1, \dots, n; j = 1, \dots, m.$$

Donde  $p_{ij}$  es el número de riesgos buenos con el atributo  $j$  de la variable  $i$  dividida por el número total de riesgos buenos (de la variable  $i$ );  $q_{ij}$  es el número de riesgos malos con el atributo  $j$  de la variable  $i$  dividido por el número total de riesgos malos (de la variable  $i$ ); y  $w_{ij} = \ln(p_{ij}/q_{ij})$ . Generalmente, cualquier variable con un valor informativo de más de 0.1 será considerado en el score-card [10].

También es importante saber que el tipo de variables que se emplean para la creación de los modelos de Credit Scoring, dependen en gran medida del tipo de cartera para la cual el modelo será construido.

Para una cartera retail (donde se encuentran las personas físicas y las empresas pequeñas), es más común el uso de variables socioeconómicas o datos de emprendimiento productivo. Mientras que para carteras corporates (que incluyen a las medianas y grandes empresas), generalmente se emplean las variables extraídas de estados contables, información cualitativa acerca de la dirección, el sector económico, proyecciones del flujo de fondos, etc.

En la elaboración de un Credit Scoring de corporates, se ha observado un patrón entre las variables seleccionadas que sugieren la importancia de 4 variables principales: ganancias, apalancamiento, tamaño y

liquidez; aunque no todas son usadas en cada estudio. En un artículo publicado por Falkenstein (2000), donde se presenta la metodología del modelo default de RiskCalc (el cual, es un software desarrollado por Moody's para la elaboración de Credit Scoring, cálculo de probabilidad default, y otros), el modelo utiliza: activos/IPC, inventarios/costo de mercaderías vendidas, pasivos/activos, crecimiento de los ingresos netos, ingresos netos/activos, prueba ácida, ganancias retenidas/activos, crecimiento en las ventas, efectivo/activos y ratio de cobertura del servicio de la deuda. El artículo también señala que, aunque la teoría recomienda utilizar ratios de apalancamiento y rentabilidad en un modelo de Scoring, la experiencia sugiere usar ratios de liquidez. En otro estudio, de Srinivasan y Kim (1987), donde se compara el desempeño de distintos modelos para deudas corporate, las variables que usan son: activo corriente/pasivo corriente, prueba ácida, patrimonio neto/deuda, logaritmo de los activos, ingresos netos/ventas, ingresos netos/activos. Y finalmente, el Z-score (Altman, 1968) utiliza: capital de trabajo/activos, ganancias retenidas/activos, EBIT/activos, valor de mercado del patrimonio neto/valor libros de la deuda y ventas/activos.

En los modelos para deudas retail, de Greene (1992) y Boyes et al. (1989) se usan variables socioeconómicas, tales como edad, estado civil, cantidad de personas a cargo, tiempo de permanencia en el domicilio actual y en el empleo actual, nivel educativo, si es propietario de la vivienda que habita, gastos mensuales promedio/ingresos mensuales promedio, tipo de ocupación, si tiene tarjeta de crédito, cuenta corriente o caja de ahorro, número de consultas en los credit bureaus y cómo está calificado en ellos.

Por otro lado, el FICO credit risk score<sup>1</sup> esencialmente usa variables asociadas al comportamiento de pagos actual y pasado, dando mayor importancia al comportamiento pasado como mejor predictor del comportamiento futuro. Los grupos de variables empleadas, junto con su incidencia en el score, son: historia de pagos (35 %), monto adeudado (30 %), largo de historia crediticia (15 %), nuevo crédito (10 %) y tipo de crédito usado (10 %).

Cabe resaltar que, a diferencia de las aplicaciones más académicas, y por las regulaciones legales (en contra de la discriminación en Estados Unidos), no utiliza variables como raza, religión, nacionalidad, sexo y estado civil. Tampoco emplea las variables de edad, ingresos, ocupación y antigüedad en el empleo, domicilio, la tasa de interés y número de consultas realizadas al buró por el deudor, por entidades financieras para ofrecer productos pre-aprobados o para monitorear a sus deudores, o por empleadores. Sin embargo, la cantidad de consultas realizadas en respuesta a solicitudes de crédito sí influye en el score. Los datos recabados para la realización de un Scoring del portafolio retail generalmente se clasifican en positivos y negativos. Principalmente, se asocia la información negativa con los atrasos e incumplimientos de pagos; mientras que la información positiva está unida a los pagos a término y la descripción de las deudas, como montos de préstamos, tasas de interés y plazos de los financiamientos.

La mayoría de los modelos Scoring para retail se enfocan en el uso de la información negativa para su desarrollo. Sin embargo, la evidencia empírica muestra que la inclusión de la información vinculada al buen comportamiento en pagos mejora sustancialmente el desempeño de los modelos. Un buen ejemplo de esto es, el trabajo realizado por Powell et

---

<sup>1</sup>Uno de los modelos desarrollados en la industria por Fair Isaac Corporation, el cual es uno de los principales desarrolladores de Credit Scoring, empleado por los tres principales burós de crédito de Estados Unidos de Norteamérica para calcular sus scores (de buró).



al. (2004) quienes, con datos de Argentina, Brasil y México, realizaron la cuantificación de la mejora en el poder predictivo de los modelos al incluir la información positiva respecto a modelos que sólo usan información negativa, y se muestra que su utilización por parte de los que otorgan créditos facilita el acceso al crédito y mejora la calidad de los portafolios de préstamos de las entidades financieras.

Por último, los modelos de Credit Scoring para micro-empresarios y PyMEs tienden a combinar información personal del titular del emprendimiento y del negocio. Uno de los primeros desarrollos fue el Small Business Scoring Solution que Fair Isaac Corporation introdujo en 1995, que fue pionero en combinar información de los principales dueños de la empresa y del negocio mismo. Dentro de los desarrollos académicos más recientes, Miller y Rojas (2005) hacen Credit Scoring de PyMEs de México y Colombia, mientras que Milena, Miller y Simbaqueba (2005) hacen lo mismo para microfinancieras de Nicaragua [9].

### 2.3.1. Construcción del Modelo

Se pueden enlistar varios pasos para la construcción de un modelo Credit Scoring, un ejemplo de ello es el de Nieto et al. (2010) [17]:

- 1.– **Conformar la base de datos.** Es la obtención de la base de datos en un formato electrónico, con que se trabajará el modelo; a partir de la información otorgada en la solicitud de crédito y el resto de las fuentes disponibles.
- 2.– **Depurar la base de datos.** En este paso se buscan los datos nulos, inconsistentes, fuera de rango o erróneos que tenga la base para corregirlos o eliminarlos. Se debe evaluar cada variable y elemento de la muestra, para determinar si se debe eliminar la variable o al elemento de la muestra, que contengan muchos valores con errores.
- 3.– **Agrupar la base de datos.** Con la base depurada se conforman intervalos de clase (atributos) para cada variable. También se debe separar la muestra, generalmente tomando, 75 % para el modelo y

25 % para la validación.

- 4.– **Seleccionar las características.** Mediante herramientas estadísticas se analiza cada variable, para saber cuales presentan mayores diferencias en las proporciones de los clientes buenos y malos, es un punto decisivo para determinar que variables se deben agregar al modelo.
- 5.– **Determinar la función de clasificación.** La clasificación de los clientes se puede realizar con cualquiera de los métodos que se mencionan a continuación.
- 6.– **Validar el modelo.** Una vez elaborada la función de clasificación, se valida tomando la parte de la muestra restante (25 %) y se corrobora que los resultados se apeguen al modelo original (back-testing).
- 7.– **Elaborar el scordcard.** En caso de que el modelo lo permita, se puede formar el scorecard con cada variable, calculando los valores con una traslación o cambios de escala. En este se tendrá el peso de cada atributo de cada variable.
- 8.– **Establecer el puntaje de corte o cut-off.** Se establece el puntaje mínimo que debe tener el solicitante para ser aceptado.

## 2.4. Modelos de Credit Scoring

Los modelos de Credit Scoring buscan producir una regla general que determine la probabilidad default de una determinada solicitud, para generar esta regla se debe analizar la relación que existe entre las características disponibles de los solicitantes. Realizando dicho análisis y con el uso de un sistema de score se puede determinar la clasificación de un solicitante. Este problema de clasificación se puede ver mediante la siguiente ecuación:

$$P = f(x_1, x_2, \dots, x_k) + \varepsilon. \quad (2.1)$$

Donde  $x_i$  son las variables explicativas,  $\varepsilon$  la perturbación aleatoria,  $f()$  la función que determina la relación entre las variables utilizadas, y  $P$  la probabilidad de que el crédito resulte en no pago. El principal objetivo de los problemas de clasificación se centra en calcular una función que permita ajustar con la mayor exactitud las observaciones de la muestra, y que los errores de predicción sean los menores posibles. Dependiendo si la función  $f()$  es conocida o no, o si se trata de un modelo paramétrico o no paramétrico.

Los modelos paramétricos se basan en una función de distribución o clasificación conocida, es decir, que la forma funcional  $f()$  se establece a priori, de modo que el problema radica en estimar los parámetros que mejor se ajusten a las observaciones de la muestra. Estos modelos son muy efectivos cuando el proceso generador de datos sigue la distribución propuesta, pero son muy sensibles a las violaciones de las hipótesis de partida cuando se utilizan muestras pequeñas.

Por otro lado, los modelos no paramétricos tratan de aproximar la función de clasificación a través del uso de formas funcionales flexibles, sin suponer ninguna estructura a priori. Por lo tanto, son más flexibles en las restricciones por lo que se considera que son más fáciles de aplicar que los modelos paramétricos, ya que permiten reconstruir la función de clasificación. Contrario al caso de los modelos paramétricos no buscan estimar los parámetros de la función, sino que mediante formas funcionales buscan aproximarse a la función objetivo [3].

A continuación se explicarán algunos de los modelos usados para la elaboración de un Credit Scoring.

### 2.4.1. Juicio Humano vs Modelos Scoring

No existe un debate sobre que método para determinar la calidad crediticia, es el más rápido, barato y consistente entre las instituciones

financieras. Aún así, hay quienes creen que, los analistas más calificados y experimentados podrían superar cualquier modelo. Debido a que los modelos se centran en un conjunto más restringido de información disponible, y se crea presumiblemente una ventaja para el analista.

Sin embargo, debemos considerar en primera instancia que los humanos tienen sesgos y limitaciones como cualquier modelo. Las “finanzas conductuales”, uno de los campos de las finanzas académicas, explica las anomalías de la teoría tradicional del “agente racional”<sup>2</sup>, mediante la cual se rige la toma de decisiones de los agentes de crédito, a través de la documentación de predisposiciones psicológicas. Entre estas predisposiciones están: la tendencia de las personas a sobreestimar la precisión de sus conocimientos (Alpert and Raiffa (1982)); el incremento de la confianza en paralelo con la importancia de la tarea; y, que es más fácil acordarse de la información relacionada con los éxitos que con los fracasos (Barber and Odean (1999)). En conclusión, se puede decir que los individuos están mal calibrados.

Un estudio que muestra evidencia empírica a favor de los modelos cuantitativos contra el juicio aplicado a préstamos, lo da Libby (1975). En su estudio, Libby preguntó a 16 agentes de crédito de bancos pequeños y a 27 de bancos grandes que juzgarán a 60 firmas y seleccionarán 30 que irían a la quiebra dentro de un período de 3 años, a partir de los estados financieros que fueron presentados. Los agentes solicitaron 5 ratios financieros para basar su decisión. Aunque estuvieron bien el 74 % en el tiempo estimado, esto fue inferior a las alternativas más simples como es la ratio de pasivos/activos.

---

<sup>2</sup>La teoría de la elección racional, o teoría de la acción racional, es utilizada para entender y modelar formalmente el comportamiento social y económico. Es la principal corriente teórica en la microeconomía y supone que el individuo o agente tiende a maximizar su utilidad-beneficio y a reducir los costos o riesgos. Los individuos prefieren más de lo bueno y menos de lo que les cause mal.

También hay que tomar en cuenta la evidencia, que muestra que mientras que los humanos son buenos para encontrar variables importantes, no son muy buenos en integrar de forma óptima fuentes de información variadas (Meehl (1954)).

Otro motivo de por qué en el pronóstico de default, los modelos cuantitativos pueden superar al juicio humano, es que usualmente el análisis no se enfoca en un objetivo estricto de fallo. Los modelos cuantitativos generalmente solo se juzgan por su potencia y calibración, mientras que el análisis de juicio humano se enfoca en proveer una explicación convincente y se basa, más en explicar las evaluaciones individuales que en un amplio rendimiento estadístico. Dado que generalmente los bancos no guardan un registro histórico para probar y calibrar los juicios de sus analistas estadísticamente, no es de extrañar que no hayan sido optimizados con objetivos estadísticos; se requiere de retroalimentación para poder mejorar el razonamiento inductivo y en la mayoría de las instituciones esa retroalimentación es dada de forma anecdótica y no estadística.

A pesar de haber fuertes razonamientos para creer que los modelos cuantitativos son invaluable para la elaboración de predicciones exactas y consistentes, no quiere decir que el juicio humano no sea útil. Al final, la ventaja competitiva recaerá en un proceso de juicio. Lo principal es que este juicio se centre en las áreas en que, de un mayor valor, en vez de en un ámbito de análisis indefinido y sin restricciones. Se debe tomar a los modelos cuantitativos como herramientas de apoyo, y no de toma de decisiones. Dar al modelo de juicio, es decir, los humanos, la última decisión, es diferente a un modelo de juicio donde sólo se depende de las decisiones tomadas por los humanos.

Se puede considerar que un proceso de toma de decisiones se base en información cuantitativa y el juicio de manera óptima si tiene las siguientes características:

Primero, la información cuantitativa debe centrarse en un número compuesto, de modo que no se presente una mezcla de ratios y factores de riesgo. Si se presentan muchos números, tienden a hacer el juicio del analista subjetivo al no poder hacer comparaciones claras entre individuos.

En segundo lugar, se debe enfocar el juicio en excepciones, donde las puntuaciones cuantitativas son extremas, pero haya circunstancias atenuantes. Mientras que el objetivo principal es un número final, en el proceso de excepciones se deben centrar en los factores que influyen el resultado. Tal es el caso de los factores que están “fuera del modelo”, que puedan afectar las perspectivas de los individuos en cuanto al futuro [7].

Por lo tanto, se puede concluir que, para hacer una óptima clasificación de las solicitudes de crédito en buenos o malos, es importante basarse en un modelo estadístico y usar el juicio de los agentes de crédito para las solicitudes que posean puntuaciones cercanas al cut-off y que ellos determinen su clasificación en buenos o malos. Aunque la mayor parte de población a clasificar dependa del modelo estadístico que se ocupe para la realización del Credit Scoring es importante de igual manera el uso del juicio humano.

### 2.4.2. Análisis Discriminante

Supongamos que tenemos un espacio de individuos  $\Omega$  constituídos por  $g$  poblaciones o clases conocidas  $C_1, \dots, C_g$  y para cada una de ellas se observa un cierto vector de interés  $X = (X_1, \dots, X_p)$ , con  $p \geq 1$ .

El análisis discriminante es un modelo paramétrico, que consiste en una técnica estadística multivariante, compuesta por un conjunto de métodos de carácter descriptivos y predictivos, que se encarga de describir los rasgos diferenciales entre las poblaciones mediante las variables

explicativas  $X_i$  ( $i = 1, \dots, p$ ). Se busca encontrar *funciones discriminantes* o *reglas de decisión*  $h = h(x_1, \dots, x_p)$  cuyos valores en las distintas clases estén lo más lejos posibles, es decir, se trata de encontrar funciones  $h$  sencillas que permitan asignar a cada individuo de la muestra a una clase concreta  $C_i$ ,  $i = 1, \dots, g$ , buscando minimizar la tasa de error en la asignación.

Dado un individuo  $\omega$ , cuya clase de procedencia se desconoce, y del cual se pueden medir las variables  $X_1, \dots, X_p$ , es decir,  $x = (x_1, \dots, x_p)$ , donde  $x_i = X_i(\omega)$ , para  $i = 1, \dots, p$ , el problema de clasificación trata de asignar a este individuo a una de las  $g$  clases, mediante el uso de funciones discriminantes construidas a partir de la muestra.

La técnica de análisis discriminante usada para el Credit Scoring es la regla discriminante lineal de Fisher, donde  $h$  es una función lineal de  $x = (x_1, \dots, x_p)$ . Sean  $\mu_1, \mu_2$  los vectores de medias de las clases  $C_1, C_2$ , respectivamente, sea  $\Sigma$  la matriz de covarianzas común para ambas clases y sea  $\omega$  el individuo a clasificar, para el cual se observa  $x = (x_1, \dots, x_p)$ .

El *criterio geométrico*, consiste en asignar al individuo  $\omega$  a la población más cercana, utilizando la distancia de Mahalanobis:

$$\delta_M^2(x, \mu_i) = (x - \mu_i)' \sigma^{-1} (x - \mu_i), \quad i = 1, 2.$$

La regla de decisión es la siguiente:

$$\begin{aligned} \omega \text{ se asigna a } C_1 & \text{ si } \delta_M^2(x, \mu_1) < \delta_M^2(x, \mu_2), \\ \omega \text{ se asigna a } C_2 & \text{ en caso contrario.} \end{aligned}$$

De la diferencia  $\delta_M^2(x, \mu_2) - \delta_M^2(x, \mu_1)$ , se tiene la función discriminante lineal

$$L(x) = (x - \frac{1}{2}(\mu_1 + \mu_2))' \sigma^{-1} (\mu_1 - \mu_2)$$

y la regla de decisión en función de ésta es:

$\omega$  se asigna a  $C_1$  si  $L(x) > 0$ ,  
 $\omega$  se asigna a  $C_2$  en caso contrario.

Esta función discriminante es el llamado discriminador lineal de Fisher.

Las hipótesis de partida del análisis discriminante son:

1. Las  $P$  variables independientes tienen una distribución normal multivariante.
2. Hay igualdad en la matriz de covarianzas de las variables independientes en cada uno de los grupos.
3. El vector de medias, las matrices de covarianzas, las probabilidades a priori, y el costo de error son magnitudes conocidas.
4. La muestra de la población es aleatoria.

Para el caso de muestras de carácter financiera las dos primeras hipótesis difícilmente son verificadas, lo cual no impide obtener buenas estimaciones, aunque en realidad no puedan considerarse óptimas [3].

### 2.4.3. Regresión Logística

Los modelos de Regresión Logística han tenido una amplia variedad de aplicaciones a lo largo del tiempo; sus primeros usos fueron en estudios en el área de biología, en los últimos años se han ocupado estos modelos en estudios en las áreas sociales y de marketing. De igual manera se ha convertido en una herramienta popular en aplicaciones en el área de negocios. Algunas aplicaciones de Credit Scoring usan estos modelos para calcular la probabilidad de incumplimiento.

La Regresión Logística se basa en un modelo logístico, en el cual se busca la relación entre dos o más variables, de forma que cada elemento  $x$  del conjunto independiente  $X$  le corresponde un único elemento  $\pi(x)$ .



Enfocando el modelo a Credit Scoring, se utiliza el modelo de regresión logística múltiple dicotómica, dada por la expresión:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}. \quad (2.2)$$

Otra forma de ver la ecuación 2.2 es:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}. \quad (2.3)$$

De estas ecuaciones se puede calcular la probabilidad de que ocurra un evento para distintos valores de las variables independientes  $x_1, \dots, x_p$ . Por lo que, la probabilidad de que el evento no ocurra esta dado por:

$$1 - \pi(x) = 1 - \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}. \quad (2.4)$$

La función logística cuenta con una función inversa llamada transformación *logit* la cual es importante para el desarrollo de la regresión ya que cuenta con propiedades de un modelo de regresión lineal, lo que permite que la probabilidad estimada se mantenga en el rango de valores 0 y 1. De la ecuación 2.3 obtenemos,

$$\begin{aligned} \pi(x) &= \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}} \\ \Rightarrow 1 + \exp^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} &= \frac{1}{\pi(x)} \\ \Rightarrow \exp^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} &= \frac{1}{\pi(x)} - 1 \\ \Rightarrow \exp^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} &= \frac{1}{\frac{1 - \pi(x)}{\pi(x)}} \\ \Rightarrow \exp^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} &= \frac{\pi(x)}{1 - \pi(x)} \\ \therefore \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p &= \ln \frac{\pi(x)}{1 - \pi(x)}. \end{aligned}$$

De ahí,

$$\text{logit}[\pi(x)] = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p. \quad (2.5)$$

El *logit* es lineal en sus parámetros, tiene su dominio (valores de  $x$ ) en el intervalo  $(-\infty, \infty)$ , y codominio (valores de  $\pi(x)$ ) en el intervalo  $(0, 1)$ , teniendo como único caso posible para que  $\frac{\pi(x)}{1-\pi(x)} > 0$ , que el numerador y el denominador sean positivos, tenemos:

1.  $\pi(x) > 0$ .
2.  $1 - \pi(x) > 0 \Rightarrow 1 > \pi(x)$ .

Con base en esto, se define la regresión añadiendo un error  $\epsilon$  y la variable  $Y$ , donde  $Y$  da a  $\pi(x)$  una interpretación de probabilidad:

$$y = \pi(x) + \epsilon = P(y|x) + \epsilon = \frac{1}{1 + e^{-x}} + \epsilon.$$

Donde  $\epsilon$  puede tomar uno de dos valores posibles. Si  $y = 1$  entonces  $\epsilon = 1 - \pi(x)$  con probabilidad  $\pi(x)$ , y si  $y = 0$  entonces  $\epsilon = -\pi(x)$  con probabilidad  $1 - \pi(x)$ , por lo que  $\epsilon$  tiene una distribución con media cero y varianza igual a  $\pi(x)[1 - \pi(x)]$ .

### Estimación de parámetros del modelo

Para estimar los parámetros  $\beta'_i$ s, los cuales representan la relación entre las variables y los diversos pesos de cada características de los clientes, se toma la función logística y la matriz  $X$  de la muestra, con  $n$  elementos de cada variable  $X_j$  y los resultados conocidos  $Y$ . Para esto se utiliza el método de máxima verosimilitud.

Con el Estimador de Máxima Verosimilitud (EMV), se puede inferir uno o más parámetros  $\beta = \beta_1, \beta_2, \dots, \beta_k$ , basándose en las  $n$  observaciones hechas  $\hat{X}$ . Se busca el valor estimado que maximiza la Función de Verosimilitud, una función conjunta de los parámetros desconocidos  $\beta$  y las observaciones  $\hat{X}$  de la muestra:

$$L(\beta) = f_1(y_1) \times f_2(y_2) \times \dots \times f_n(y_n) = \prod_{i=1}^n f_i(y_i).$$

Para obtener la estimación de los parámetros  $\hat{\beta}$ , la maximización debe cumplir la siguiente desigualdad:  $L(\hat{\beta}) \geq L(\beta)$ . Para realizar esta estimación comúnmente se hace uso de la función logaritmo,

$$L(\beta) = \ln(f_1(y_1)) \times \ln(f_2(y_2)) \times \cdots \times \ln(f_n(y_n)) = \sum_{i=1}^n \ln(f_i(y_i)).$$

Para la función logística  $\pi(X)$ , se consideran a las variables  $Y$  como un evento Bernoulli donde cada elemento observado tiene la posibilidad de ser un éxito o un fracaso, de ahí tenemos que:

$$f_i(y_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i},$$

$$L(\beta) = \ln(l(\hat{x}_i; \beta)) = \sum_{i=1}^n [y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i))]. \quad (2.6)$$

Para encontrar el valor del vector  $\beta$  que maximiza  $L(\beta)$ , se deriva  $L(\beta)$  con respecto a  $\beta$  y se igualan las derivadas a 0, para poder despejar el conjuntos de parámetros. Primero se deriva con respecto a  $\pi(\hat{x}_i; \beta)$ :

$$\frac{\partial l(\hat{x}_i; \beta)}{\partial \pi(\hat{x}_i; \beta)} = \sum_{i=1}^n \left[ \frac{y_i}{\pi(\hat{x}_i; \beta)} - \frac{1 - y_i}{1 - \pi(\hat{x}_i; \beta)} \right] \frac{\partial \pi(\hat{x}_i; \beta)}{\partial \beta}. \quad (2.7)$$

Después se deriva con respecto a la combinación lineal  $X \cdot \beta$ , y posteriormente con respecto a cualquier coeficiente  $\beta_i$ :

$$\frac{\partial \pi(\hat{x}_i; \beta)}{\partial X \cdot \beta} \frac{\partial X \cdot \beta}{\partial \beta_i} = \frac{\partial \pi(\hat{x}_i; \beta)}{\partial \beta}$$

$$\begin{aligned} \frac{\partial \pi(\hat{x}_i; \beta)}{\partial X \cdot \beta} &= \frac{-e^{-(X \cdot \beta)}(-1)}{(1 + e^{-(X \cdot \beta)})^2} = \frac{e^{-(X \cdot \beta)}}{(1 + e^{-(X \cdot \beta)})^2} \\ &= \frac{e^{-(X \cdot \beta)}}{(1 + e^{-(X \cdot \beta)})} \frac{1}{(1 + e^{-(X \cdot \beta)})} \\ &= \left[ 1 - \frac{1}{(1 + e^{-(X \cdot \beta)})} \right] \frac{1}{(1 + e^{-(X \cdot \beta)})} \\ &= [1 - \pi(\hat{x}_i; \beta)] \pi(\hat{x}_i; \beta). \end{aligned} \quad (2.8)$$

Por lo que, para cada  $\beta_i \neq \beta_0$ , con  $i = 0, 1, \dots, k$ , tenemos que  $\frac{\partial X \cdot \beta}{\partial \beta_i} = X_i$  y  $\frac{\partial X \cdot \beta}{\partial \beta_0} = 1$ . De 2.7 y 2.8 se tienen las derivadas parciales

para  $\beta_0$  y los  $k$  coeficientes restantes  $\beta_i$ , estas ecuaciones se igualan a 0 para obtener los estimadores, y se tiene:

$$\begin{aligned} \frac{\partial l(\hat{X}; \beta)}{\partial \beta_i} &= \sum_{i=1}^n \left[ \frac{y_i}{\pi(\hat{x}_i; \beta)} - \frac{1 - y_i}{1 - \pi(\hat{x}_i; \beta)} \right] [1 - \pi(\hat{x}_i; \beta)] \pi(\hat{x}_i; \beta) \hat{x}_i \\ &= \sum_{i=1}^n [y_i [1 - \pi(\hat{x}_i; \beta)] \hat{x}_i - (1 - y_i) \pi(\hat{x}_i; \beta) \hat{x}_i] \\ &= \sum_{i=1}^n [y_i - \pi(\hat{x}_i; \beta)] \hat{x}_i. \end{aligned} \quad (2.9)$$

$$\frac{\partial l(\hat{X}; \beta)}{\partial \beta_0} = \sum_{i=1}^n [y_i - \pi(\hat{x}_i; \beta)]. \quad (2.10)$$

Al tratarse de ecuaciones no lineales las cuales son más complejas a mayor cantidad de elementos en la muestra y características a evaluar, se requiere de un proceso robusto que implica una gran cantidad de operaciones para su solución, en la actualidad existen diversos paquetes estadísticos que facilitan el cálculo de los parámetros.

Es importante verificar la significancia estadística de los parámetros, existen diversos estadísticos para esto, siendo el estadístico de Wald uno de los más usados.

El **estadístico de Wald** contrasta la hipótesis de que un coeficiente es distinto de 0, y sigue una distribución Normal Estándar [12].

$$H_0 : \beta_i = 0 \quad vs \quad H_1 : \beta_i \neq 0$$

Este estadístico se obtiene dividiendo el valor del estimado del coeficiente ( $\hat{\beta}_i$ ) entre su correspondiente error estándar  $\hat{\sigma}(\beta_i)$ .

$$Wald = \frac{\hat{\beta}_i}{\hat{\sigma}(\beta_i)}. \quad (2.11)$$

Para evaluarlo se calcula una Ji-cuadrada con 1 grado de libertad ( $\chi_{\alpha,1}^2$ ), que representa la variable a evaluar y se rechaza  $H_0$  cuando el

valor del coeficiente de Wald es mayor. La obtención de significación indica que dicho coeficiente es diferente de 0, y por lo tanto, el coeficiente se debe conservar en el modelo. En modelos con errores estándar grandes, el estadístico de Wald puede proporcionar falsas ausencias de significación (es decir, se incrementa el error tipo II).

### Interpretación de los parametros

Para interpretar los parámetros escribimos la ecuación en términos de odd o momio, el odd de un evento se define como la razón de probabilidad de que ocurra ese evento a la probabilidad de que no ocurra:

$$odd = \frac{P(y = 1|x)}{P(y = 0|x)} = \frac{P(y = 1|x)}{1 - P(y = 1|x)} = \frac{\pi(\hat{X}; \hat{\beta})}{1 - \pi(\hat{X}; \hat{\beta})}. \quad (2.12)$$

Estos cocientes, cuentan el número de veces que será más probable que ocurra un éxito del evento correspondiente con cada variable  $i$ .

La razón de momios (Odds Ratios) sirve para evaluar cuantitativamente el impacto de cambiar el valor de una variable; estos factores de cambio requieren de dos modelos para el análisis de cada variable  $i$ , uno que contenga todas las variables ( $modA$ ) y otro que no tenga la variable  $i$  ( $modB$ ). La razón de momios de ambos modelos es:

$$OR = \frac{\frac{\pi(\hat{X}; \hat{\beta}_{modA})}{1 - \pi(\hat{X}; \hat{\beta}_{modA})}}{\frac{\pi(\hat{X}; \hat{\beta}_{modB})}{1 - \pi(\hat{X}; \hat{\beta}_{modB})}}.$$

Tomando el logaritmo a la razón de momios tenemos:

$$\begin{aligned}
\ln(OR) &= \ln \left[ \frac{\frac{\pi(\hat{X}; \hat{\beta}_{modA})}{1 - \pi(\hat{X}; \hat{\beta}_{modA})}}{\frac{\pi(\hat{X}; \hat{\beta}_{modB})}{1 - \pi(\hat{X}; \hat{\beta}_{modB})}} \right] \\
&= \ln \left[ \frac{\pi(\hat{X}; \hat{\beta}_{modA})}{1 - \pi(\hat{X}; \hat{\beta}_{modA})} \right] - \ln \left[ \frac{\pi(\hat{X}; \hat{\beta}_{modB})}{1 - \pi(\hat{X}; \hat{\beta}_{modB})} \right] \\
&= \sum_{j=1}^k \hat{\beta}_j \cdot \hat{x}_j - \sum_{j=1, j \neq i}^k \hat{\beta}_j \cdot \hat{x}_j = \hat{\beta}_i \cdot \hat{x}_i.
\end{aligned}$$

Ahora eliminamos el logaritmo aplicando la función exponencial en ambos lados:

$$OR = e^{\hat{\beta}_i \cdot \hat{x}_i}.$$

Entonces, tenemos el factor de cambio  $e^{\hat{\beta}_i \cdot \hat{x}_i}$ , para utilizarlo se ocupan intervalos de  $u$  unidades para saber el efecto que tendrá la probabilidad de éxito, se calcula  $e^{\hat{\beta}_i \cdot u}$  [12].

#### 2.4.4. Árboles de Regresión y Clasificación (CART)

Los árboles de clasificación y regresión son una técnica de clasificación binaria, que reúne tanto características de los modelos univariante y multivariantes, su metodología usa datos históricos para construir árboles de clasificación o regresión con el fin de clasificar o predecir nuevos datos. Este modelo no paramétrico fue desarrollado en los años 80 por Breiman, Freidman, Olshen y Stone.

Estos árboles pueden usar variables numéricas y/o categóricas. Algunas de las ventajas que presenta este modelo son: su robustez a valores atípicos, la invarianza en la estructura de sus árboles de clasificación o de regresión a transformaciones monótonas de las variables independientes, y su interpretabilidad. El algoritmo de particionamiento recursivo es ocupado en el proceso para la construcción de un árbol de decisión y es clave para el método CART.

Sea  $Y$  la variable respuesta y sean  $p$  las variables predictoras  $x_1, x_2, \dots, x_p$  donde las  $x_i$  son fijas. El problema consiste en encontrar la relación entre ellas, para poder predecir  $Y$  basado en los valores de  $x$ . Se quiere estimar la probabilidad condicional de la variable  $Y$

$$P[Y = y | X_1 = x_1, X_2 = x_2, \dots, X_p = x_p]$$

o una forma funcional de la probabilidad como la esperanza condicional

$$E[Y = y | X_1, X_2, \dots, X_p].$$

Un árbol esta conformado por:

- **Nodo raíz** es un nodo único, representa el inicio del árbol como se observa en la Figura 2.1 (a). Este nodo contiene una muestra de individuos, el árbol crece, ya que, se agregan nuevos nodos llamados hijos o descendientes. Estos individuos constituyen una muestra de aprendizaje, la cual puede ser la muestra total o una parte de ésta.
  
- **Nodos internos** son subconjuntos de la muestra de aprendizaje, contiene un test sobre algún valor de las características de la muestra. En la Figura 2.1 (a) está representado por el círculo del segundo renglón.
  
- **Nodos terminales u hojas** son aquellos que no tiene descendencia, representados por cajas en la Figura 2.1 (a) con los números I, II, III.

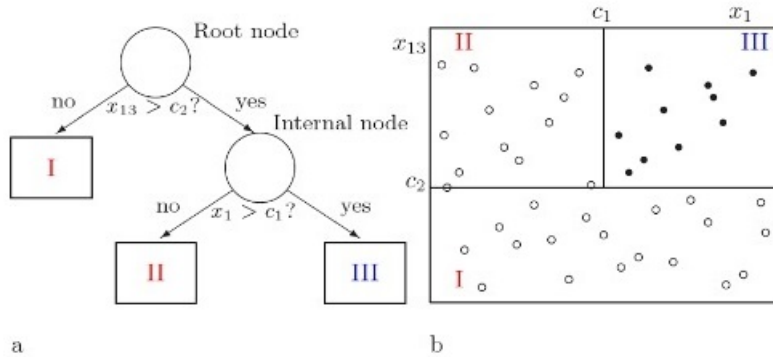


Figura 2.1: (a) Se presenta la estructura de un árbol, que está constituido por un nodo raíz, nodos internos y nodos de salida. En la parte (b), están representadas las salidas del árbol, con una clasificación homogénea.

El objetivo principal de la partición recursiva es acabar con nodos terminales homogéneos, como se muestra en la Figura 2.1 (b) donde cada nodo terminal (I, II, III) solo contiene puntos o círculos. La homogeneidad completa en los nodos terminales es rara vez alcanzada en el análisis de datos reales. Por lo tanto, el objetivo realista de la partición recursiva es hacer que las variables de salida de los nodos terminales sean lo más homogéneas posibles. Una manera sencilla de medir la homogeneidad del nodo se llama impureza del nodo y la fórmula es:

$$IPN = \frac{NSCCN}{NTSN}.$$

Donde IPN es la Impureza del nodo, NSCCN es el Número de sujetos que cumplen la característica en el nodo y NTSN el Número total de sujetos en el nodo. Entre más cercana sea la razón a 0 o 1, más homogéneo es el nodo.

El método de árboles CART sigue 3 pasos:

- Construcción de un árbol saturado con el algoritmo de partición recursiva.



- Selección del tamaño del árbol.
- Clasificación de nuevos datos usando el árbol construido.

Para construir un árbol saturado con el algoritmo de partición recursiva, primero se estudian las variables explicativas para determinar para cuál de ellas y para que valores es posible aumentar la homogeneidad del nodo. Para la división es necesario tener en cuenta la impureza del nodo, una de las fórmulas más comunes para medir la impureza, es la de entropía de Shannon:

$$i(t) = - \sum_j^J p(j/t) \log(j/t),$$

y para respuestas categóricas es más común el uso de una extensión del índice de Gini:

$$i(t) = 1 - \sum_{j=1}^J p(j/t).$$

Donde  $p(t/j)$  es la proporción de la clase  $j$  con respecto al nodo  $t$ . El algoritmo optará por la división que mejore la impureza, tomando en consideración ambos nodos descendientes. Se sigue dividiendo los nodos hasta que resulte imposible mejorar realizando más divisiones o el nodo obtenido tenga el valor mínimo, en este punto el árbol se considera saturado.

Posteriormente se inicia la etapa de poda, la cual consiste en encontrar el subárbol que sea más predictivo y menos vulnerable al ruido en los datos. Para encontrar este subárbol se busca que los nodos terminales sean homogéneos (lo más posible), ya que la calidad de un árbol está en la calidad de sus nodos terminales. Por tanto, para un árbol  $\mathcal{T}$  se define:

$$R(\mathcal{T}) = \sum_{\tau \in \tilde{\mathcal{T}}} P[\tau] r[\tau]. \quad (2.13)$$

Donde  $\tilde{\mathcal{T}}$  es el conjunto de nodos terminales de  $\mathcal{T}$ ,  $P[\tau]$  es la probabilidad de que un sujeto caiga en el nodo terminal  $\tau$  y  $r(\tau)$  es una medida

de calidad del nodo  $\tau$  la cual es similar a la suma de cuadrados de los residuales en regresión lineal. El propósito en esta etapa es seleccionar el mejor subárbol,  $\mathcal{T}^*$ , de un árbol saturado inicial,  $\mathcal{T}_0$ , tal que  $R(\mathcal{T})$  sea mínimo. Una medida de calidad del nodo  $r(\tau)$  puede ser la medida de la impureza, aunque es usual que se tome el costo de mala clasificación debido a que los árboles de clasificación tienen respuestas binarias, el costo de mala clasificación se define como:

$$r(\tau) = \sum_{i=1}^l c(j|i)P[Y = 1|\tau], \quad \text{con } j, i = 1, \dots, l. \quad (2.14)$$

Donde  $Y$  una variable dicotómica con valores 0 y 1, y sea  $c(i|j)$  el costo de mala clasificación para un sujeto de la clase  $j$  que es clasificado en la clase  $i$ . Si  $i = j$ , entonces la clasificación es correcta y el costo debería ser cero, esto es  $c(i|i) = 0$ . Para estimar el costo de mala clasificación incondicional del nodo  $\tau$  se multiplica  $r(\tau)$  por  $P[\tau]$ , que es:

$$R(\tau) = P[\tau]r(\tau). \quad (2.15)$$

Remplazando la ecuación (2.15) en (2.13), se tiene el costo de mala clasificación del árbol  $\tau$ :

$$R(\mathcal{T}) = \sum_{\tau \in \tilde{\mathcal{T}}} R(\tau). \quad (2.16)$$

En la práctica es difícil medir el costo relativo de  $c(j|i)$  para  $i \neq j$ , por lo que no se le puede asignar un costo de mala clasificación a cada nodo antes de aumentar cualquier árbol. Sin embargo, en la práctica se ha visto que el uso de una función de impureza como la entropía da árboles útiles con tamaños de muestra razonables.

Uno de los principales problemas que se presentan a la hora de construir los árboles es la cantidad de nodos, una gran cantidad de nodos puede llevar a problemas de sobreajuste. La medida de calidad del árbol debe tomar en cuenta tanto la calidad de los nodos terminales así como el tamaño del árbol. Tener en cuenta solo el costo de mala clasificación

puede llevar a árboles muy grandes.

El costo de complejidad para el árbol  $\tau$ , se define como:

$$R_\alpha(\mathcal{T}) = R(\mathcal{T}) + \alpha|\tilde{\mathcal{T}}|, \quad (2.17)$$

donde  $\alpha(\geq 0)$  es el parámetro de complejidad y  $|\tilde{\mathcal{T}}|$  es el número de nodos terminales en  $\mathcal{T}$  llamado *complejidad* del árbol  $\mathcal{T}$ . A diferencia de  $R(\mathcal{T})$ , el costo de complejidad  $R_\alpha(\mathcal{T})$  penaliza a un árbol grande.

La medida de costo complejidad permite construir una secuencia de subárboles óptimos anidados desde cualquier árbol  $\mathcal{T}$  dado. Se construye una secuencia de subárboles anidados para un árbol saturado  $\mathcal{T}$ , minimizando el costo complejidad  $R_\alpha(\mathcal{T})$ , y se selecciona como subárbol final el que tenga el costo de mala clasificación más pequeño.

Si se dispone de una muestra de prueba, estimar  $R(\mathcal{T})$  es sencillo para cualquier subárbol  $\mathcal{T}$ , ya que sólo se necesita aplicar los subárboles a la muestra de prueba y se escoge el mejor valor de  $\alpha$ . En el caso en el que no se tiene una muestra se pueden crear muestras artificiales utilizando el proceso de validación cruzada y así estimar  $R(\mathcal{T})$  [25].

### 2.4.5. Redes Neuronales

Las redes neuronales son un conjunto de modelos comúnmente usados en la minería de datos, debido a su potencia predictiva, flexibilidad y la facilidad de uso. La definición para una red neuronal varía de acuerdo al área en la que se utiliza. De forma general, se puede decir que una red neuronal es una máquina diseñada para modelar la forma en la que el cerebro realiza una actividad específica o una función de interés, esta red está compuesta por un conjunto de neuronas artificiales interconectadas.

El entrenamiento de una red neuronal se hace mediante la repetición de pruebas en varias sub-muestras de la muestra hasta que la red alcanza

una mínima diferencia entre el resultado deseado y el obtenido.

Entre las principales ventajas que presentan las redes neuronales se encuentra la no linealidad, la adaptabilidad que tiene la red a pequeños cambios en el medio que la rodea, y que se trata de un modelo no paramétrico.

Por lo que, se puede decir que las redes neuronales plantean exigencias mínimas en la estructura y supuestos del modelo. Ya que no es necesario suponer a priori determinadas relaciones entre las variables dependientes e independientes. La forma en que las variables se relacionan se determina durante el proceso de aprendizaje. Si una relación lineal entre las variables dependientes e independientes es adecuada, los resultados de la red neuronal deben aproximarse lo máximo posible a los del modelo de una regresión lineal. Mientras que, si una relación no lineal es más adecuada, entonces la red neuronal se aproximará automáticamente a la estructura “correcta” del modelo.

El inconveniente que se da con esta flexibilidad es que las ponderaciones sinápticas de una red neuronal no son fáciles de interpretar, por lo que es mejor usar un modelo estadístico más tradicional si se quiere explicar el proceso subyacente que produce las relaciones entre las variables. En las redes neuronales utilizadas en las aplicaciones predictivas, como las redes de perceptrones multicapa (MLP) o las de función de base radial (RBF), los resultados pronosticados por el modelo se pueden comparar con los valores conocidos de las variables de destino [23].

### **Estructura MLP**

La red de perceptrones multicapa (MLP) es una función de predictores (denominados también entradas o variables independientes) que minimizan el error de predicción de las variables de destino (también denominadas salidas).

La arquitectura del Perceptrón Multicapa se denomina arquitectura feedforward, debido que las conexiones de la red fluyen unidimensionalmente desde la capa de entrada hasta la capa de salida sin ciclos de retroalimentación. Esta arquitectura está compuesta por:

- La capa de entrada, la cual contiene los predictores.
- Una o más capas ocultas, que contienen nodos (o unidades) no observables. El valor de cada unidad oculta es una función de los predictores; la forma exacta de la función depende, por un lado, del tipo de red y, por otro lado, de especificaciones dadas.
- La capa de salida compuesta por las respuestas. Cada unidad de salida es una función de las entradas ocultas. Así como en las capas ocultas, la forma exacta de la función depende, del tipo de red y de las especificaciones dadas.

### Error de entropía cruzada

Las redes neuronales son generalmente entrenadas con el objetivo de minimizar una función de error; para los problemas de clasificaciones binarias una elección apropiada para esta función, es el de error de entropía cruzada:

$$E = - \sum_{i=1}^n y'_i \log y_i + (1 + y'_i) \log (1 - y_i). \quad (2.18)$$

Donde  $y_i$  es el resultado de la neurona de salida,  $y'_i$  el resultado esperado u objetivo.

Por lo tanto, se dice que el error de entropía cruzada es la suma de la relación de las diferencias entre el valor real y el valor esperado u objetivo de todas las salidas de la red.

En la propagación hacia atrás (Back propagation) este error se minimiza mediante la actualización interactiva de los pesos de las neuronas de entrenamiento [16].

### Perceptrón Multicapa

El procedimiento Perceptrón Multicapa (MLP) genera un modelo predictivo para una o más variables dependientes (de destino) basada en los valores de las variables predictoras. Un MLP tiene 3 características distintivas:

- El modelo de cada neurona en la red incluye una función de activación no lineal, la cual relaciona la suma ponderada de unidades de una capa, con los valores de unidad en la capa correcta. Una función comúnmente usada es la sigmoideal definida como:

$$y_j = \frac{1}{1 + \exp(-v_j)}. \quad (2.19)$$

Donde  $v_j$  es la suma ponderada de todas las entradas sinápticas más el sesgo de la neurona  $j$ , y  $y_j$  es la salida de la neurona.

- La red contiene una o más capas ocultas, distintas a la capa de entrada y la capa de salida, estas capas ocultas son las que permiten a la red “aprender” mediante la extracción de información significativa de la capa de entrada.
- La red muestra un alto grado de conectividad, determinada por la sinápsis de la red.

Es importante suponer que la función de activación sea diferenciable para la aplicación del algoritmo de aprendizaje de propagación hacia atrás (Back Propagation). Las neuronas sigmoideas presentan una mayor pendiente cuando trabajan más cerca de su rango medio; y una pendiente mínima con tendencias a 0 cuando las neuronas trabajan cerca de sus límites. Debido a esto se busca que las neuronas sigmoideas se entrenen con salidas destino entre 0.1 y 0.9, y no en 0 y 1 donde la derivada se aproxima a cero y por lo tanto, el cambio en los pesos del algoritmo de propagación hacia atrás será mínimo, haciendo que el aprendizaje sea más lento. Por lo que para el algoritmo de propagación hacia atrás se recomienda el uso de la función de activación tangente hiperbólica, dado por:

$$\varphi(v) = \frac{1 - \exp(-v)}{1 + \exp(-v)}. \quad (2.20)$$

Esta función permite que el algoritmo de Back Propagation aprenda más rápido en términos de iteraciones requeridas, debido a que el rango de salida de la función va de -1 a 1, por lo que la función hiperbólica se considera simétrica ya que cumple con  $\varphi(-v) = -\varphi(v)$ .

Una importante recomendación de este método es que la función de activación pueda variar para esperar distintos resultados. De esta forma cada problema busca encontrar la arquitectura neuronal más conveniente.

Cuando la red neuronal intenta obtener modelos de funciones, es decir, una aproximación de funciones lineales o no lineales es necesario que en la capa de salida sean neuronas lineales y usar funciones de activación de tangente hiperbólica en las demás capas. Ya que las neuronas de salida lineal permiten acceder a rangos de salida distintos a cero y uno. En las capas ocultas, las neuronas no lineales permiten a la red aprender hasta encontrar la función de aproximación.

En esta metodología, también hay que tomar en cuenta el problema de la saturación incorrecta de neuronas, lo cual sucede cuando el algoritmo de entrenamiento se estanca en un mínimo local o superficial, de esta forma las neuronas se saturan antes de tiempo provocando que trabajen cerca de los límites de la función que es donde la pendiente se aproxima a cero por lo que los cambios en los pesos serán mínimos.

Una medida para evitar los problemas de saturación, es iniciar los pesos uniformemente distribuidos dentro de un rango pequeño, de esta manera se puede evitar que la función de activación se dispare a zonas de activación. Otra recomendación es mantener un número bajo de neuronas ocultas, ya que las neuronas extras pueden acarrear consecuencias

graves como que la red aprenda idiosincrasias en los patrones evitando un conocimiento acertado, por lo que se debe procurar trabajar a las neuronas en la zona media donde se considera que la región es lineal.

Hay dos formas para entrenar una red neuronal, con respecto a la presentación de los datos a la red y los ajustes en los pesos sinápticos, son:

- Entrenamiento de patrón por patrón, a la red se le presenta un patrón e inmediatamente después se realiza la propagación hacia atrás ajustando los pesos. Este entrenamiento es útil debido a que el ajuste en los pesos sucede de forma aleatoria ya que con cada patrón surge un nuevo ajuste de pesos, de esta forma se evita que el algoritmo caiga en un mínimo local y el uso de la memoria se reduce, ya que no almacena cambios.
- Entrenamiento conjunto de datos o Batch, los datos se presentan a la red y el cálculo de la modificación de los pesos es almacenado hasta que todas las muestras sean presentadas a la red, luego se modifican los pesos con la suma de las modificaciones almacenadas. Este algoritmo es bueno ya que estima de manera más precisa el vector gradiente y tiene una mayor eficiencia computacional. Este método generalmente es más adecuado para los problemas de clasificación y detección de patrones.

Se le conoce como epoch a cada presentación completa a la red neuronal del conjunto de entrenamiento. Una recomendación, a tomar en cuenta sin importar cuál de los métodos de aprendizajes mencionados anteriormente se ocupe, es que el orden de las muestras sea aleatorizado de epoch en epoch para aumentar la velocidad de la convergencia. La velocidad en la convergencia es sumamente importante, ya que de ésta forma el tiempo desperdiciado en el entrenamiento de la red disminuye, y también se asegura que la red trabaje correctamente con el conjunto de entrenamiento mostrado. Es importante resaltar que, aunque el algoritmo converja a un valor pequeño no implica una buena generalización.



Las redes MLP entrenadas apropiadamente ofrecen respuestas razonables cuando se ponen a prueba con entradas que no han visto. Esta propiedad de generalización hace posible que el entrenamiento de la red con un conjunto representativo de entradas y salidas esperadas, obtener buenos resultados sin necesidad de entrenar a la red con todos los pares posibles de entradas y salidas.

No obstante, la generalización no siempre se logra, aunque una reducción en el error es posible. Los factores que influyen en la obtención de una buena generalización son, el tamaño y la eficiencia del conjunto de datos de entrenamiento, la arquitectura de la red y la complejidad física del problema. El último punto no se puede controlar, pero los dos primeros sí.

Después de fijar la arquitectura de la red, la cual se determina de acuerdo a la complejidad física del problema, utilizando la ecuación (2.21), propuesta por Haykin, se puede determinar el tamaño del conjunto de datos por:

$$N \geq \frac{W}{\epsilon}. \quad (2.21)$$

Donde  $W$  es el número de pesos de la red y  $\epsilon$  es la fracción de error permitido, regularmente se establece en .01, de esta forma se puede asegurar que el número de muestras permitirá a nuestra red generalizar correctamente [11].



## Capítulo 3

# Problemática

La República de China (o ROC por sus siglas en inglés) está compuesta por un conjunto de islas, cuya área es aproximadamente  $36,000 \text{ km}^2$ . Está situada al suroeste de Japón, al norte de las Filipinas y alrededor de 180 km de la costa sureste de China. Las principales islas que abarca el territorio de la ROC son Taiwán, Kinmen, Matsu y Penghu. La moneda nacional es el Nuevo Dólar Taiwanés, cuyo código es NT\$ o TWD.

A partir de la Segunda Guerra Mundial la economía en Taiwán experimentó un crecimiento, del mismo modo se dió la industrialización en el país. En 1987 el país comenzó un camino hacia la democratización política y la liberación económica, como resultado del levantamiento de la ley marcial. Esto hizo que, Taiwán rápidamente se transformará de un fabricante de bienes básicos a una economía industrial, volviéndose uno de los países más competitivos en el sector mundial de tecnologías de la información y las comunicaciones. En el ámbito político, también se ha desarrollado como una democracia multipartidista. Hoy en día, Taiwán disfruta de uno de los más altos estándares de vida en Asia y también está posicionado como uno de los primeros en términos de liberalización económica, libertad de prensa, salud, derechos humanos y libertad política.

El mercado financiero de la ROC se divide en:

- Mercado de Dinero para las transacciones de corto período y préstamos interbancarios.
- Mercado de Capitales, que incluye el mercado primario, el mercado secundario, las securitizaciones y servicios auxiliares.
- Mercado de Moneda Extranjera, donde hay dos subdivisiones, una para el público en general y el interbancario.

Para el año 2002, existían un total de 404 instituciones captadoras de ahorros, entre las cuales se encuentran: Bancos, Cooperativas de Ahorro y Crédito, Departamento de Crédito de las Asociaciones de Agricultores y Pescadores, Compañías de Inversiones, Sistema Postal, Compañías de Seguros, Compañías Bursátiles Financieras, Compañías de Securitizaciones y Compañías Bursátiles de Inversiones.

Antes del año 2004, la supervisión de las instituciones financieras se encontraba a cargo del Ministerio de Finanzas (MOFA), sus tres agencias (las cuales son el Buró de Asuntos Monetarios (BOMA), la Comisión de Futuros y Securitizaciones) y el Departamento de Seguros. Después, a inicios de Julio de 2004 se estableció la Comisión de Supervisión Financiera (FSC por sus siglas en inglés) dependiente del Poder Ejecutivo, la cual es actualmente la única agencia reguladora y supervisora de las instituciones financieras. Cuya misión es mantener la estabilidad financiera, acelerar la internacionalización y desregulación del sistema, facilitar un ambiente saludable y propicio para las inversiones e incrementar la confianza del mercado, educar y proteger a los inversores.

En esta sección se verán algunos antecedentes en la economía de Taiwan que lo llevó a la crisis de tarjetas de crédito que se vivió durante 2005 y finales de 2006. Así como la descripción de la base de datos con la que se trabajó [5].

### 3.1. Crisis Financiera de Taiwan

A principios de 1990 el gobierno taiwanés permitió la formación de nuevos bancos. Estos nuevos bancos se dedicaron a prestar grandes cantidades de dinero a compañías de Bienes Raíces para expandir su negocio y aumentar sus ganancias. Después de unos años el mercado de Bienes Raíces se saturó y las ganancias comenzaron a disminuir.

Entonces los bancos buscaron otro negocio, el de tarjetas de crédito y tarjetas de pago. Para expandir este nuevo negocio los bancos lanzaron campañas publicitarias para animar a la población a solicitar tarjetas de crédito para consumir. Redujeron los requisitos para obtener créditos con el fin de obtener más clientes entre la población joven. A pesar de que esta población generalmente no contaba con suficientes ingresos, los bancos continuaron otorgándoles créditos.

En Febrero de 2006 las deudas de las tarjetas de crédito y tarjetas de pago alcanzaron los \$268 billones de dólares. Más de medio millón de personas no fueron capaces de pagar sus préstamos y se convirtieron en “esclavos de tarjetas de créditos”, término inventado en Taiwan para referirse a las personas que solo podían pagar el mínimo en sus tarjetas de crédito cada mes.

Esta situación llegó a ser un gran problema para la sociedad, algunos deudores y sus familias cometieron suicidios, otros se volvieron indigentes, algunos no pudieron pagar la educación de sus hijos e incluso algunos de los “esclavos de tarjetas de crédito” vendían drogas para pagar a los bancos.

En el 2005, para prevenir el incremento en las tarjetas de crédito o “esclavos de tarjetas de crédito”, la Comisión de Supervisión Financiera (FSC) emitió ordenes a los bancos para que modificaran sus requerimientos de tarjetas de crédito y tarjetas de pago. Algunos de estos

nuevos requerimientos eran contar con un mayor ingreso, se prohibió el uso de campañas publicitarias inadecuadas, así como los comportamientos de cobro inadecuados y las tasas de interés compuesto [22].

## 3.2. Descripción de la Base de Datos

Se trabajó con una base de datos que contiene información del segundo y tercer trimestre de 2005, de un importante banco de Taiwán. La base cuenta con 30,000 observaciones de las cuales 6,636 (22.1 %) son “clientes morosos”. Esta base cuenta con una variable respuesta binaria (No moroso=0, Moroso=1). La base cuenta con 23 variables explicativas:

- X1: **Monto del crédito** dado, esto incluye ambos, el crédito individual consumido y el crédito de su familia.
- X2: **Género** (Hombre=1; Mujer=2).
- X3: **Educación** (1= Posgrado; 2=Universidad; 3=Preparatoria; 0,4,5,6=Otro).
- X4: **Estado civil** (1=Casado; 2=Soltero; 3=Divorciado; 0=Otros).
- X5: **Edad** (en años).
- X6-X11: **Historial de pagos pasados**. Se tiene el record mensual de los pagos pasados (de Abril a Septiembre, 2005) de la siguiente manera: X6= status de pago de Septiembre de 2005; ...; X11= status de pago de Abril de 2005. La medida de la escala para el status de pago es:
  - -2= No hubo consumo; -1= Pago completo; 0= Uso de crédito revolvente; 1= Pago retrasado por un mes; 2= Pago retrasado por dos meses; ...; 8= Pago retrasado por 8 meses; 9= Pago retrasado por 9 meses o más.
- X12-X17: **Monto del estado de cuenta**. X12= Monto del estado de cuenta en Septiembre 2005; ...; X17= Monto del estado de cuenta en Abril de 2005.

- X18-X23: **Monto de pago previo.** X18= Monto pagado en Septiembre 2005; X19= Monto pagado en Agosto 2005; ...; X23= Monto pagado en Abril 2005.

### 3.2.1. Descripción de las Variables

En esta sección se presentan las descripciones de las variables que contiene la base de datos, así como cierta información relevante.

El **monto de crédito**, se refiere tanto al monto de crédito otorgado al cliente como al crédito de su familia (suplementario). La mayoría de los montos de crédito otorgados se encuentran en los rangos de NT<sup>1</sup>\$10,001 a NT\$208,000, estos créditos son el 66.28 % de las observaciones. En el gráfico de la Figura 3.1 se puede observar que los créditos mayores a NT\$505,000 son escasos (representan el 0.69 %).

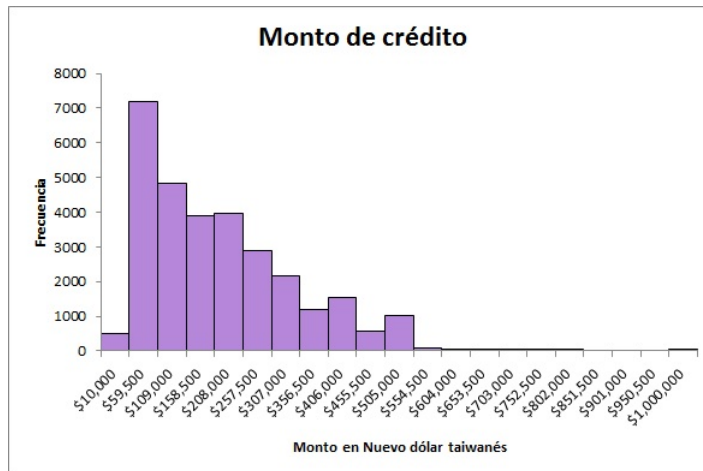


Figura 3.1: Histograma de frecuencia del Monto de crédito

A continuación se muestran las Tablas I y II de frecuencias para las variables cualitativas nominales: **Género**, **Educación** y **Estado Civil**. Se observa, que en su mayoría los clientes son mujeres (60.4 %); también se observa que casi en su totalidad los clientes cuentan con una

<sup>1</sup>NT : Nuevo Dólar Taiwanés

educación mayor a la media superior (82.1%), y que el porcentaje de clientes solteros es mayor (53.2%).

Tabla I: Frecuencias de Género

	Frecuencia	Porcentaje
Hombre	11,888	39.6 %
Mujer	18,112	60.4 %
Total	30,000	100 %

Tabla II: Frecuencia del Nivel Educativo

	Frecuencia	Porcentaje
Posgrado	10,585	35.3 %
Licenciatura	14,030	46.8 %
Preparatoria	4,917	16.4 %
Otro	468	1.5 %
Total	30,000	100 %

Tabla III: Frecuencias del Estado Civil

	Frecuencias	Porcentaje
Casado	13,659	45.5 %
Soltero	15,964	53.2 %
Divorciado	323	1.1 %
Otro	54	.2 %
Total	30,000	100 %

En la Figura 3.2 se muestra el siguiente histograma de frecuencia de la variable **edad**, donde se puede ver que la mayor parte de las observaciones se encuentran en el rango de 22 a 46 años (86.11%).



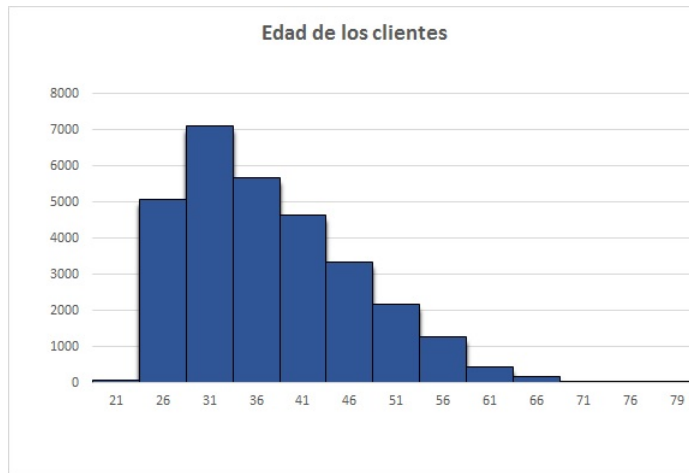


Figura 3.2: Histograma de frecuencia de la variable Edad

Para las variables del **Historial de pagos pasados** se tiene una tabla de frecuencias, que se encuentra en el anexo, en estas tablas se detallan de Abril a Septiembre de 2005, si no se realizó consumo, si el crédito fue pagado, si se hizo uso de crédito revolvente o si se tiene un retraso en el pago (desde un mes hasta 8 meses de retraso se registran). De esa tabla, se puede notar que de los meses observados (de Abril de 2005 a Septiembre 2005) la mayoría de los clientes hacen uso del crédito revolvente (de 49.1 % a 56.5 %).

Para las variables de **Estado de cuenta** tenemos los Estadísticos Descriptivos en la Tabla IV; esta variable muestra el saldo del crédito mensual, de Abril a Septiembre de 2005, de los clientes. De donde se puede observar que el saldo promedio en los estados de cuenta de todos los meses observados está alrededor de los \$45,000 (en la moneda Nuevo Dólar Taiwanés), por otro lado, en los valores mínimos aparecen cuentas observadas con saldo a favor (con signo negativo) y el crédito más alto se registró en Julio de 2005 por \$1,664,089 (en la moneda Nuevo Dólar Taiwanés).

Tabla IV: Descriptivos del Monto del estado de cuenta

	Mínimo	Máximo	Promedio	Desviación
Sep 05	-\$165,580	\$964,511	\$51,223.33	\$73,635.86
Ago 05	-\$69,777	\$983,931	\$49,179.08	\$71,173.77
Jul 05	-\$157,264	\$1,664,089	\$47,013.16	\$69,349.39
Jun 05	-\$170,000	\$891586	\$43,262.95	\$64,332.86
Mayo 05	-\$81,334	\$927171	\$40,311.40	\$60,797.16
Abril 05	-\$339,603	\$961664	\$38,871.76	\$59,554.11

En las variables de **Monto de pago previo** mostrado en la Tabla V, se ve que en promedio cada mes los clientes realizaron pagos alrededor de \$5,000.

Tabla V: Descriptivos del Monto previo pagado

	Máximo	Promedio	Desviación
Monto pagado en Sep 05	\$873,552	\$5,663.58	\$16,563.28
Monto pagado en Ago 05	\$1,684,259	\$5,921.16	\$23,040.87
Monto pagado en Jul 05	\$896,040	\$5,225.68	\$17,606.96
Monto pagado en Jun 05	\$621,000	\$4,826.08	\$15,666.16
Monto pagado en Mayo 05	\$426,529	\$4,799.39	\$15,278.31
Monto pagado en Abril 05	\$528,666	\$5,215.50	\$17,777.47

Para la variable respuesta, si el cliente es o no moroso, es decir, si incurre o no en el no pago. De la Tabla VI se nota que el 77.9 de los clientes son no morosos.

Tabla VI: Frecuencias de la Variable Respuesta

	Frecuencias	Porcentaje
Moroso	6,636	22.1 %
No Moroso	23,364	77.9 %
Total	30,000	100 %

## Capítulo 4

# Una aplicación del Credit Scoring

Para la aplicación de la técnica de Credit Scoring se usó una base de datos (descrita en el capítulo anterior) de un banco de Taiwán con información de Abril a Septiembre de 2005. La base de datos fue obtenida de UCI Machine Learning Repository [14], otorgada por PhD. I-Cheng Yeh [24].

Esta base de datos cuenta con la información de 30,000 individuos, de los cuales 6 fueron excluidos debido a valores inconsistentes entre las variables, la base tiene 23 variables explicativas y una variable respuesta.

A partir de esta base de datos se busca construir un modelo de Credit Scoring para la clasificación de clientes. Se usó la paquetería de SPSS [21] en la base de datos, para crear una red neuronal MLP, para la clasificación de los clientes. En el Anexo B se muestran los pasos que se siguieron para la construcción de la red en SPSS.

Se realizaron diversas pruebas para elegir una red neuronal, se cam-

bio el orden de las variables <sup>1</sup>, se eliminaron variables y se probaron diferentes tipos de entrenamiento.

En la tabla 4.1 se muestran algunos de los resultados que se obtuvieron, ya que varias redes tuvieron el mismo porcentaje global correcto, se eligió la red que presentó el menor error de entropía cruzada.

Tabla 4.1: Comparación de resultados.

	Error de entropía cruzada	Porcentaje global correcto		
		Entrenamiento	Pruebas	Reserva
En el orden en que aparecen	7,375.824	81.2 %	81.4 %	80.6 %
Quitando (4) variables	7,245.680	82.0 %	82.0 %	81.2 %
De mayor a menor importancia	7,233.872	81.7 %	81.7 %	81.4 %
Quitando (4) variables	7,239.881	81.8 %	81.8 %	81.4 %
De menor a mayor importancia	7,380.174	81.8 %	81.5 %	81.4 %
Quitando (4) variables	7,236.68	82.1 %	81.8 %	81.4 %

A continuación se muestran los resultados obtenidos por la red neuronal elegida, en la cual se ordenaron las variables independientes de mayor a menor importancia, de acuerdo a los resultados obtenidos previamente ordenando las variables en el orden en que aparecen.

## 4.1. Resultados

Se asignaron 16,663 (55.6 %) individuos a la muestra de entrenamiento, 4,233 (14.1 %) individuos a la muestra de prueba y 9,098 (30.3 %) a la muestra de reserva, como se muestra en la Tabla 4.2.

---

<sup>1</sup>Ya que en la construcción de redes con SPSS el orden de las variables modifica el resultado.

Tabla 4.2: Resumen de procesamiento de casos.

		N	Porcentaje
Muestra	Entrenamiento	16,663	55.6 %
	Prueba	4,233	14.1 %
	Reserva	9,098	30.3 %
Válido		29,994	100 %
Excluido		6	
Total		30,000	

La tabla de información sobre la red, Tabla 4.3, es útil para asegurar que las especificaciones de la red son correctas. En esta tabla se detallan las especificaciones de la red neuronal, el número de unidades en la capa de entrada es el número de covariables más el número de factores de nivel; se crea una unidad adicional por cada categoría de las variables. Se utilizaron 10 unidades en la capa oculta de la red.

Tabla 4.3: Información de la red

Capa de entrada	Factores	9	
	Covariables	14	
	Número de unidades <sup>a</sup>	88	
	Método de cambio de escala para las covariables	Estandarizado	
Capas ocultas	Número de capas ocultas	1	
	Número de unidades en la capa oculta 1 <sup>a</sup>	10	
	Función de activación	Tangente Hiperbólica	
Capa de salida	Variables dependientes	1	Tipo de cliente
	Número de unidades	2	
	Función de activación	Softmax	
	Función de error	Entropía cruzada	

a. Se excluye la unidad de sesgo.

La Tabla 4.4 muestra el resumen de los resultados obtenidos de entrenar y probar la red, y aplicar la red a la muestra reserva (9,098 clientes). Se observa que el porcentaje de pronósticos incorrectos es aproximadamente igual en las muestras de entrenamiento, prueba y reserva. Al ser

la regla de parada que se produjera un paso(s) consecutivo(s) sin disminución de error sugiere que no se produjo un sobreentrenamiento en la red, esta regla de parada nos indica que el entrenamiento de la red fue bueno.

Tabla 4.4: Resumen del modelo

Entrenamiento	Error de entropía cruzada	7,233.872
	Porcentaje de pronósticos incorrectos	18.3 %
	Regla de parada utilizada	1 paso(s) consecutivo(s) sin disminución del error <sup>a</sup>
Prueba	Error de entropía cruzada	1,833.707
	Porcentaje de pronósticos incorrectos	18.3 %
Reserva	Porcentaje de pronósticos incorrectos	18.6 %

Variable dependiente: Tipo de cliente

a. Los cálculos de error se basan en la muestra de comprobación.

En la Tabla 4.5, se detallan los porcentajes de clasificación correcta en las tres muestras de la base, para cada tipo de cliente (moroso o no moroso). Se observa que el porcentaje correcto global es similar para las tres muestras. En la muestra reserva para la clasificación de clientes no morosos tiene un porcentaje correcto del 95.6 %, para los clientes moroso un 32.9 % y el porcentaje correcto global es del 81.4 %. Por lo que se puede decir que en general la red tiene un buen porcentaje correcto de pronósticos.

En la Figura 4.1, se muestran gráficos de caja para cada categoría de la variable respuesta, muestran como se distribuyen las probabilidades pronosticadas para la clasificación de los clientes<sup>2</sup>.

<sup>2</sup>a partir del 0.5 de probabilidad se considera a un cliente moroso

Tabla 4.5: Clasificación

Muestra	Observado	Pronosticado		Porcentaje correcto
		No moroso	Moroso	
Entrenamiento	No moroso	12,423	613	95.3 %
	Moroso	2,428	1,199	33.1 %
	Porcentaje global	89.1 %	10.9 %	81.7 %
Pruebas	No moroso	3,145	148	95.5 %
	Moroso	627	313	33.3 %
	Porcentaje global	89.1 %	10.9 %	81.7 %
Reserva	No moroso	6,725	308	95.6 %
	Moroso	1,386	679	32.9 %
	Porcentaje global	89.2 %	10.8 %	81.4 %

Variable dependiente: Tipo de cliente.

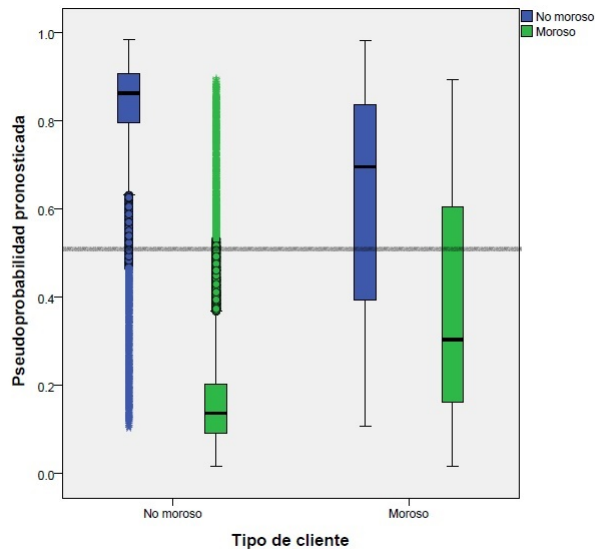


Figura 4.1: Gráfico Pseudoprobabilidad Pronosticada.

A continuación se muestra la curva ROC, Figura 4.3, que representa la capacidad discriminativa de la red para clasificar morosos como morosos y no morosos como no morosos. Un parámetro para evaluar la bondad de una prueba diagnóstica que produce resultados continuos es el área bajo la curva (AUC, por sus siglas en inglés) la cual se muestra

en la Figura 4.2, esto es, que si se elige un cliente no moroso hay un 77 % de probabilidad de se clasifique correctamente como no moroso.

**Área bajo la curva**

		Área
Tipo de cliente	No moroso	.770
	Moroso	.770

Figura 4.2: Área bajo la curva.

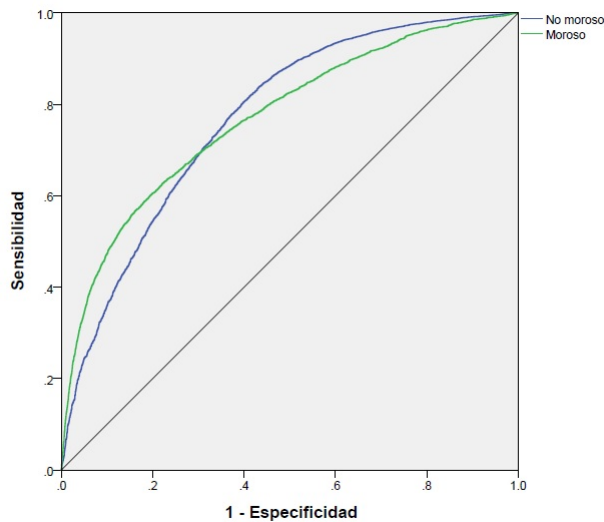


Figura 4.3: Curva ROC.

Por último se muestra el gráfico de la importancia normalizada, en la Figura (4.4), de cada variable utilizada en la red neuronal, es decir, cuánto cambia el valor predictivo de la red para diferentes valores de la variable independiente. Podemos observar que las variables predictivas más importantes son las que nos hablan sobre los comportamientos financieros de los clientes, estas son, pagos previos, monto del estado de cuenta, historial de pagos pasados y límite de crédito. Siendo el status de



pago del mes de julio la variable más importante. Por otro lado, las variables con el menor poder predictivo son género, edad, nivel de estudios y estado civil.

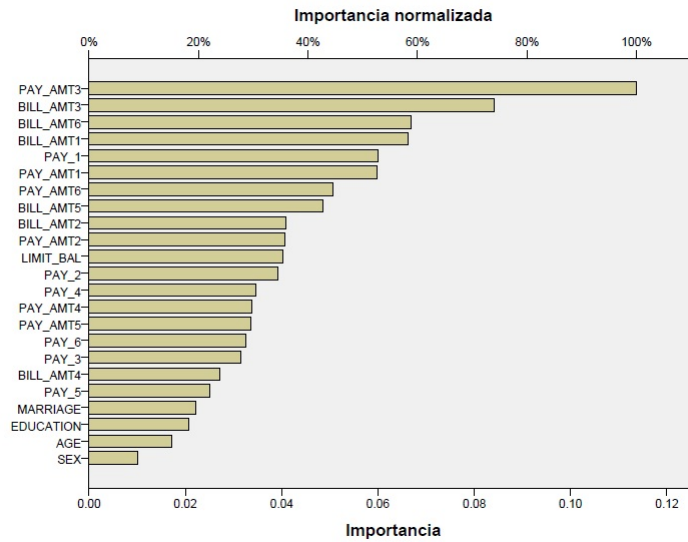


Figura 4.4: Importancia normalizada.



## Capítulo 5

# Conclusiones

En la actualidad existen empresas especializadas en el desarrollo de modelos de Credit Scoring para su uso comercial. Estos modelos son de utilidad para empresas que prefieren pagar por su uso a desarrollar sus propios modelos, también son usados por algunas instituciones gubernamentales. Al ser estos modelos construidos para uso comercial son efectivos para una buena clasificación de los clientes.

Por otro lado, para desarrollar un Credit Scoring se debe adecuar a la base de datos con la que se plantea trabajar, el tipo de cliente y la información con la que se cuenta de los solicitantes. Es necesario que las instituciones determinen previamente la definición de clientes buenos y malos, ya sea con base en la experiencia previa o tomando en cuenta la experiencia de fuentes externas y así limitar los parámetros para definir a los tipos de clientes. Con base a estos parámetros, las instituciones solicitarán la información conveniente a sus clientes, para poder aplicar un Credit Scoring.

Un aspecto importante al realizar un Credit Scoring, es separar la muestra para realizar un backtesting y comprobar si la clasificación obtenida es adecuada. En los Credit Scoring que se ocupan en la rea-

lidad es necesario que se hagan backtesting de forma recurrente para actualizar correctamente los ajustes a las variables que se utilizan.

En el caso práctico que se realizó, la base de datos se separó en 3 muestras; la de entrenamiento con la que la red neuronal crea la primera prueba, la muestra de prueba en donde la primera red obtenida vuelve a entrenarse para mejorar y en la muestra de reserva se pone a prueba la red obtenida.

En este trabajo se estudió y revisó la metodología del Credit Scoring, y se realizó una aplicación práctica con una base de datos de un banco de Taiwán que se consiguió en línea. Al tratarse de una base de datos grande se optó por usar una técnica de minería de datos, tal como son las redes neuronales, que han demostrado su poder predictivo en el desarrollo de estos sistemas. Además, las redes neuronales cuentan con la ventaja de ser un método flexible y de fácil aplicación.

Se han desarrollado diversas arquitecturas de redes neuronales, para la aplicación se utilizó el modelo de redes neuronales Perceptron Multicapa (MLP). Debido a que se ha visto que su uso es eficaz en el desarrollo de Credit Scoring; y además existen diversos software mediante los cuales se puede aplicar.

Mediante el uso de SPSS se hicieron varias pruebas: cambiando el orden de las variables, usando diferentes tipos de entrenamiento y eliminando variables. Ya que el cambio en estos parámetros altera los resultados obtenidos. Se eligió la red neuronal con menor error de entropía y mayor porcentaje de pronósticos correctos.

La red neuronal MLP final que se obtuvo tiene un porcentaje global correcto de 81.4%, por lo que a pesar de que el porcentaje correcto de clientes no morosos es menor al 50% aún se puede considerar que la red obtenida tiene buenos resultados. Sin embargo, es importante resaltar

que en los Credit Scoring que se ocupan en casos reales es necesario que el porcentaje de clasificación incorrecta de clientes morosos sea el mínimo posible, ya que se considera que este tipo de mala clasificación representa el mayor costo, ya que el que los clientes no paguen representa pérdidas.

La red neuronal que se obtuvo tiene un punto de corte de 0.5, el cual es establecido por default en el programa, para considerar a un cliente moroso o no moroso. A partir de los gráficos obtenidos se observó que si se reduce este punto de corte, por ejemplo a 0.4, el porcentaje correcto de clientes morosos aumentaría en mayor proporción a la disminución en el porcentaje correcto de clientes no morosos.

De los resultados obtenidos del caso práctico se observó que las variables de mayor importancia son aquellas que dan información financiera del solicitante, las más importante fueron el monto previo pagado en julio de 2005 y el monto del estado de cuenta en julio; las variables con menor importancia son las que dan información demográfica de los solicitantes, están son: estado civil, nivel de educación, edad y el género.

Esto concuerda con lo que se ha visto en trabajos recientes donde las variables que hablan sobre los pagos que han realizados los clientes son más usadas debido a su poder predictivo en comparación con las variables demográficas que cada vez son menos usadas en los modelos de Credit Scoring. También se vio que con el método de redes neuronales se obtuvo un buen porcentaje de pronosticos correctos, por lo que como se ha visto en diversos trabajos esta técnica es buena en la elaboración de Credit Scoring.



# Apéndice A

## Estadística

### A.1. Distancia de Mahalanobis

La Distancia de Mahalanobis es una medida útil para determinar la similitud entre dos variables aleatorias multidimensionales.

De manera formal, se define a la distancia de Mahalanobis entre dos variables aleatorias, con la misma distribución de probabilidad,  $\bar{x}$  y  $\bar{y}$  y con matriz de covarianza  $\sigma$ , como:

$$d_m(\bar{x}, \bar{y}) = \sqrt{(\bar{x} - \bar{y})^T \sigma^{-1} (\bar{x} - \bar{y})}$$

Otra definición es, considerando muestras multivariantes de  $k$  poblaciones con vectores de medias  $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$  y matriz de covarianzas común  $S$ . La distancia al cuadrado de Mahalanobis entre las poblaciones  $i, j$  es:

$$M^2(i, j) = (\bar{X}_i - \bar{X}_j)' S^{-1} (\bar{X}_i - \bar{X}_j).$$

Esta medida de distancia entre variables, sirve para conocer no solo la distancia entre variables sino también su correlación [4].

La distancia de Mahalanobis cumple con las propiedades necesarias para ser una distancia:

- Semipositividad: esto es, que si dos puntos tienen las mismas coordenadas entonces la distancia entre ellos es cero, y si tienen distintas coordenadas entonces su distancia es positiva, nunca negativa.

$$d(a, b) \geq 0, \quad \forall a, b \in X.$$

$$d(a, b) = 0, \quad \text{si } a = b.$$

- Simetricidad

$$d(a, b) = d(b, a), \quad \forall a, b \in X.$$

- Desigualdad triangular

$$d(a, b) \leq d(a, c) + d(c, b), \quad \forall a, b, c \in X$$

[6].

## A.2. Distribución Normal Multivariada

Debido a que muchas técnicas multivariadas suponen que sus datos se generan de una distribución normal multivariada, esta distribución de probabilidad multivariada es considerada de las más importantes, se define de la siguiente manera:

**Definición.** Sea  $X' = (X_1, \dots, X_p)$  un vector aleatorio. Se dice que  $X$  tiene una distribución normal multivariada si su función de densidad está dada por:

$$f(x) = (2\pi)^{-\frac{p}{2}} \exp\left\{-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right\} \prod_{i=1}^p I_{(-\infty, \infty)}(x_i),$$

donde  $\mu \in R^p$  y  $\Sigma$  es una matriz definida positiva [13].

## A.3. Pruebas Aplicadas en el Análisis Discriminante

La prueba **M de Box** compara la igualdad de las matrices de covarianzas en los grupos, contrasta la hipótesis nula de que las matrices



de covarianzas poblacionales son iguales,  $H_0 : \Sigma_1 = \Sigma_2$ . Este estadístico tiene la forma:

$$M = (n - g) \log |S| - \sum_{j=1}^g (n_j - 1) \log |S_j|,$$

donde  $S$  es la matriz de varianzas-covarianzas combinadas,  $S_j$  es la matriz de varianzas-covarianzas del grupo  $j$ -ésimo,  $n$  el número total de casos y  $g$  el número de grupos.

Es importante notar que este test es sensible a la falta de normalidad multivariante, esto es, que matrices iguales pueden parecer como significativamente diferentes si no hay normalidad. De igual manera si las muestras son grandes la prueba pierde efectividad (rechazar la hipótesis nula es más fácil).

Esta prueba es comúnmente usada para probar el supuesto de igualdad entre las matrices de covarianzas de los grupos. Aunque en la práctica la técnica de análisis discriminante funciona bien aunque no se cumpla esta restricción.

La prueba  $\lambda$  de **Wilks** es la razón entre la suma de cuadrados dentro de los grupos y la suma de cuadrados totales, es decir, que mide las desviaciones dentro de cada grupo respecto a las desviaciones totales<sup>1</sup> (sin distinguir grupos). Toma valores entre 0 y 1, si la media de todos los grupos observados son iguales entonces el Lambda es uno, esto es, que entre más cerca de uno esté la Lambda menor es el poder discriminante de la variable considerada y entre más cercana a 0 es mayor su poder discriminante.

Este estadístico permite contrastar la hipótesis nula,  $H_0$  : las medias de los grupos son iguales, es decir, no existe diferencia entre los grupos. Si el p-valor asociado al estadístico es inferior al nivel de significancia

---

<sup>1</sup>Desviación Total = Desviación dentro de los grupos + Desviación entre grupos

(normalmente 0.05) se rechaza la hipótesis de igualdad entre los grupos, concluyendo que la información aportada por las respectivas variables es estadísticamente significativa. El valor de este estadístico es el que se usa generalmente para seleccionar las variables que se usan en el modelo [18].

## Apéndice B

# Tablas de Frecuencias de la Base de Datos

En las siguientes tablas, Tabla B.1 y Tabla B.2 se usan las abreviaciones: Crédito R para Crédito Revolvente; R un mes para Retraso de un mes, R 2 meses para Retraso de 2 meses y así sucesivamente hasta R 8 meses que es Retraso de 8 meses en el pago.

Tabla B.1: Frecuencias en el Historial de Pagos Pasados  
Segundo trimestre del 2005

	Jun 05		Mayo 05		Abril 05	
	Frec	%	Frec	%	Frec	%
Sin consumo	4,348	14.5 %	4,546	15.2 %	4,895	16.3 %
Pagado	5,687	19.0 %	5,539	18.5 %	5,740	19.1 %
Crédito R	16,455	54.9 %	16,947	56.5 %	16,286	54.3 %
R un mes	2	0 %	0	0 %	0	0 %
R 2 meses	3,159	10.5 %	2,626	8.8 %	2,766	9.2 %
R 3 meses	180	0.6 %	178	0.6 %	184	0.6 %
R 4 meses	69	0.2 %	84	0.3 %	49	0.2 %
R 5 meses	35	0.1 %	17	0.1 %	13	0 %
R 6 meses	5	0 %	4	0 %	19	0.1 %
R 7 meses	58	0.2 %	58	0.2 %	46	0.2 %
R 8 meses	2	0 %	1	0 %	2	0 %

Tabla B.2: Frecuencias en el Historial de Pagos Pasados  
Tercer Trimestre del 2005

	Sep 05		Ago 05		Jul 05	
	Frec	%	Frec	%	Frec	%
Sin consumo	2,759	9.2 %	3,782	12.6 %	4,085	13.6 %
Pagado	5,686	19.0 %	6,050	20.2 %	5,938	19.8 %
Crédito R	14,737	49.1 %	15,730	52.4 %	15,764	52.5 %
R un mes	3,688	12.3 %	28	0.1 %	4	0 %
R 2 meses	2,667	8.9 %	3,927	13.1 %	3,819	12.7 %
R 3 meses	322	1.1 %	326	1.1 %	240	0.8 %
R 4 meses	76	0.3 %	99	0.3 %	76	0.3 %
R 5 meses	26	0.1 %	25	0.1 %	21	0.1 %
R 6 meses	11	0 %	12	0 %	23	0.1 %
R 7 meses	9	0 %	20	0.1 %	27	0.1 %
R 8 meses	19	0.1 %	1	0 %	3	0 %

## Apéndice C

# Red Neuronal en SPSS

Primero fijamos la semilla para generar números aleatorios, de esta forma los resultados de las corridas no variarán, y se pueden replicar los resultados obtenidos. Primero seleccionamos el menú “Transformar”, luego la opción “Generadores de números aleatorios...”, después activamos la opción “Establecer punto de partida”, “Valor fijo” y luego se establece la semilla con la que se empezará a generar los números aleatorios, como se muestra en la Figura C.1; en este caso se fijo el número de partida en 9191972.

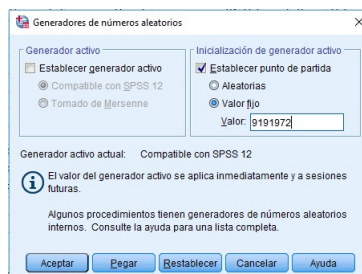


Figura C.1: Valor fijo para generar números aleatorios.

Se crea una partición para separar la muestra en 3 submuestras: entrenamiento, prueba y reserva. En este caso primero se construye una variable llamada “Particion”, con una distribución Bernoulli (0.7) modi-

ficada para que genere resultados 1 y -1, luego se realiza una modificación para que se generen resultados de 1 y 0 a partir de los resultados mayores a 0. Como se muestran en la Figura C.2 y posteriormente la modificación en la Figura C.3. De esta forma se tiene  $100 * (0.7 * 0.8) = 56\%$  con valor de 1, aproximadamente 14% con valor de 0 y 30% con valor de -1, a la muestra de entrenamiento se le asignan los casos con valor de 1, para la muestra de prueba con valor de 0 y para la de reserva se ocupan los casos con valor de -1.

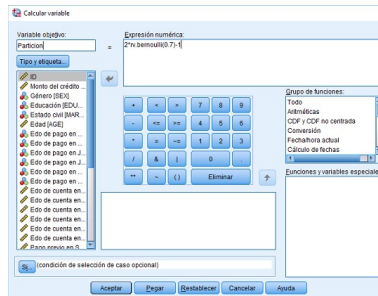


Figura C.2: Crear la variable “Particion”.

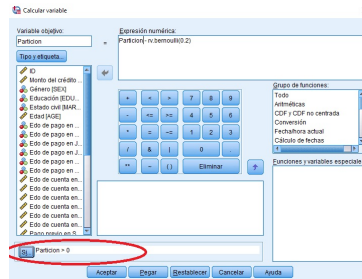


Figura C.3: Modificación a la variable “Particion”.

Para crear la red neuronal MLP, seleccionamos el menú “Analizar”, luego la opción “Redes neuronales” y luego “Perceptrón multicapa...”. De la ventana que se abre primero en la pestaña “Variables”, en “Variables dependientes” se selecciona la variable dependiente en este caso es la variable “Tipo de cliente”, en la sección de “Factores” se seleccionan

las variables explicativas que son de tipo nominal y en “Covariables” se seleccionan las variables que son de tipo escala. Como se muestra en la Figura C.4.

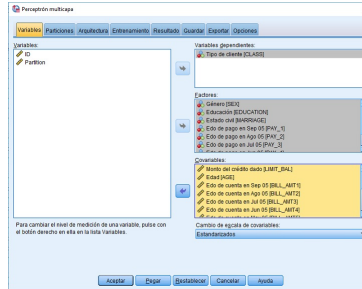


Figura C.4: Selección Variables.

En la pestaña “Particiones” se selecciona la variables que se creó previamente, “Particion” activando la opción “Utilizar variable de partición para asignar los casos”; como se muestra en la Figura C.5. También se puede “Asignar aleatoriamente los casos según el número relativo de casos” pero en este caso el “Número relativo” asignado a cada partición es un número entero.

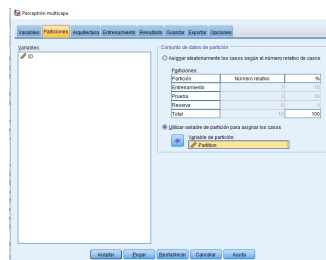


Figura C.5: Particiones.

En la pestaña “Arquitectura” (Figura C.6) se tiene dos opciones “Selección automática de arquitectura” o “Arquitectura personalizada” en esta opción se pueden personalizar los aspectos de la red para las capas ocultas y la capa de salida. La selección automática de la arquitectura construye una red con una capa oculta y calcula el número de unidades

en la capa oculta, y utiliza las funciones de activación por defecto de las capas ocultas y de salida. En la “Arquitectura personalizada” se da un control de cada aspecto de la red:

- Capa oculta. Se seleccionan una o dos capas ocultas y la función de activación, la cual vincula las sumas ponderadas de unidades en una capa a los valores de unidades en la capa siguiente.
- Capa salida. Contiene los valores de la variable dependiente y se selecciona la función de activación. En caso necesario se hace un “Cambio de escala de variables dependientes de escala”.

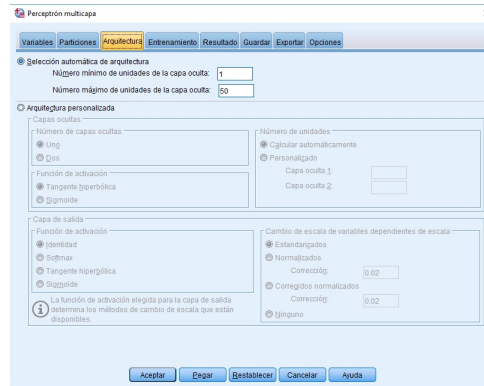


Figura C.6: Arquitectura de la Red.

En la pestaña de “Entrenamiento” (Figura C.7) se puede especificar el tipo de entrenamiento y el algoritmo de optimización determina las opciones de entrenamiento disponibles. El tipo de entrenamiento es el que determina como la red procesa las epoch (o registros) que genera. El tipo de entrenamiento se puede seleccionar de acuerdo al tamaño de la base con la que se trabaja:

- Lote ( o de Batch). Es útil para base de datos pequeños, ya que actualiza los pesos sinápticos de la red continuamente hasta que se cumpla alguna regla de parada, es decir, que puede necesitar



muchos “pases” de datos. La ventaja de este tipo de entrenamiento es que reduce directamente el error total.

- En línea. Este tipo de entrenamiento es adecuado para base de datos “más grandes” y con predictores asociados, ya que utiliza la información de un registro a la vez.
- Mini lote. La formación de mini-lotes es una “combinación” entre el entrenamiento por lotes y en línea, y puede ser mejor para conjuntos de datos de “tamaño mediano”.

El “Algoritmo de Optimización” es el método que se utiliza para estimar los pesos sinápticos, puede ser:

- Gradiente conjugado de escalado. Los supuestos para el uso de este método sólo se aplican a los tipos de entrenamiento por lotes, por lo que no está disponible para el entrenamiento en línea o por mini lotes.
- Pendiente de gradiente. Este método puede ser usado para los tres tipos de entrenamiento.

Las “Opciones de Entrenamiento” permiten afinar el algoritmo de optimización, generalmente no necesitan cambios a menos que la red tenga problemas con la estimación

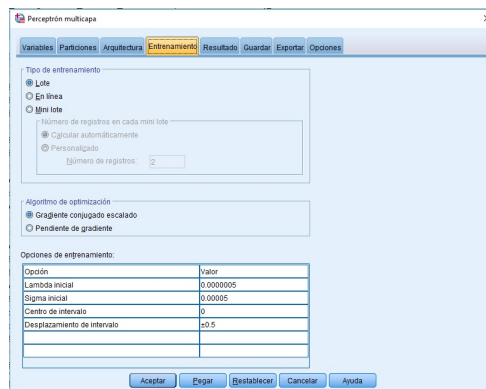


Figura C.7: Entrenamiento.

En la pestaña “Resultado” (Figura C.8), se puede seleccionar que salidas queremos observar. En la sección “Estructura de red” se muestra información resumida de la red:

- Descripción. Muestra las variables dependientes, el número de unidades de entrada y salida, el número de capas y unidades ocultas y las funciones de activación.
  
- Diagrama. Muestra el diagrama de la red neuronal de forma gráfica.
  
- Ponderaciones sinápticas. Muestra las ponderaciones sinápticas solo de la muestra de entrenamiento, hay que tener en cuenta que el número de ponderaciones puede llegar a ser grande y que usualmente no se utilizan para interpretar la red.

De la sección “Rendimiento de la red” obtenemos los resultados que sirven para determinar si el modelo es “bueno”.

El “Resumen de procesamiento de casos” despliega una tabla con el número de casos incluidos y excluidos en el análisis, por cada muestra que se uso y en total.

Y la opción “Análisis de importancia de la variable independiente” despliega un gráfico y una tabla que muestra la importancia y la importancia normalizada de cada predictor, este es obtenido mediante un análisis de sensibilidad y se basa en las muestras combinadas de entrenamiento y de prueba (en caso de no haber muestra de prueba solo en la muestra de entrenamiento).

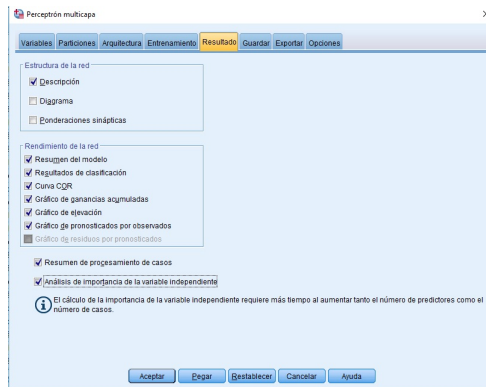


Figura C.8: Resultado.

La pestaña “Guardar” (Figura C.9) se usa para guardar las predicciones como variables en la base de datos, se pueden guardar las categorías o valores pronosticados o la pseudoprobabilidad de cada variable dependiente; los nombres de estas variables se pueden personalizar o crear automáticamente.

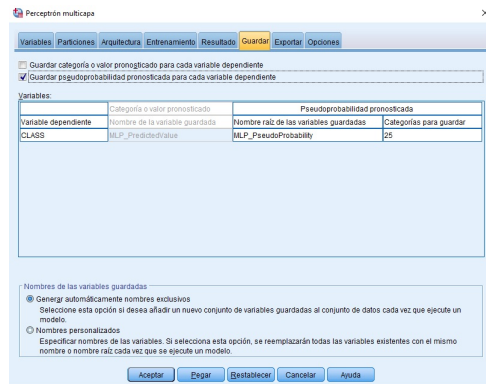


Figura C.9: Guardar.

La pestaña “Exportar” se usa para guardar las estimaciones de las ponderaciones sinápticas para cada variable dependiente en un archivo XML (PMML). Con este archivo se puede aplicar la información del modelo a otros archivos de datos.

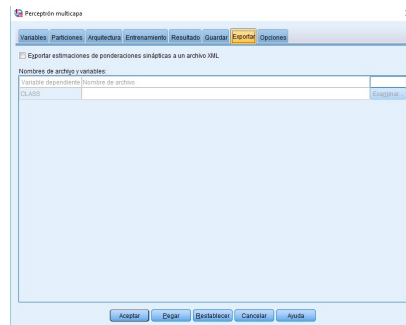


Figura C.10: Exportar.

La pestaña “Opciones” se puede seleccionar que hacer con los valores perdidos, la regla de parada y el número máximo de casos que se almacenan:

- Valores perdidos del usuario. Sirve para determinar si los valores perdidos serán tratados como válidos entre los factores y variables dependientes categóricas.
- Reglas de parada. Se selecciona la regla de parada, esta es la que determina cuándo dejar de entrenar la red neuronal.
- Número máximo de casos que se almacenarán en memoria. Se debe especificar un número mayor a 1.

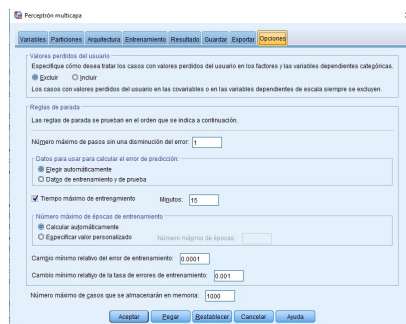


Figura C.11: Opciones.

Cuando se hayan seleccionado todas las especificaciones necesarias para la red neuronal MLP que se quiere crear se selecciona “Aceptar” [21].



# Apéndice D

## Glosario

**Activo.** Es un recurso controlado por una entidad, identificado, cuantificado en términos monetarios, del que se esperan beneficios económicos futuros, derivado de operaciones ocurridas en el pasado, que han afectado económicamente a dicha entidad.

**Activo circulante.** Que comprende los activos más líquidos, incluye el efectivo y los activos que se convertirán en efectivo en el transcurso de un año a partir de la fecha del balance general.

**Apalancamiento.** El grado de endeudamiento en su estructura de capital de la entidad.

**Capital de trabajo o capital neto de trabajo.** Es la diferencia entre los totales del activo y pasivo circulante.

**Deuda.** Las empresas emiten (venden) documentos conocidos como deuda (contratos de préstamo) o acciones de capital (certificados accionarios).

**EBIT.** Una cifra de particular importancia es la de utilidades antes

de intereses e impuestos (UAI, o EBIT, siglas de earnings before interest and taxes) que resume las utilidades antes de impuestos y costos de financiamiento.

Utilidades antes de intereses pagados e impuestos. UAI se conoce por lo general como “utilidad de operación” en el estado de resultados y es la utilidad antes de partidas extraordinarias y partidas de operación canceladas o extraordinarias. Para calcular UAI, los gastos de operación se restan del total de los ingresos de operación.

Inventarios. Se compone de las materias primas que se usarán en la producción, el trabajo en proceso y los artículos terminados.

Liquidez. Se refiere a la facilidad y rapidez con que los activos se pueden convertir en efectivo (sin pérdida significativa de valor).

Pasivo. Es una obligación presente de la entidad, virtualmente ineludible, identificada, cuantificada, en términos monetarios y que representa una disminución futura de beneficios económicos, derivada de operaciones ocurridas en el pasado, que han afectado económicamente a dicha entidad.

Patrimonio neto. El capital contable de los accionistas (o patrimonio neto de la empresa) representa la diferencia entre el valor de los activos y la deuda de la empresa.

Prueba ácida (o Razón rápida). Es una prueba más rigurosa de liquidez ya que indica la solvencia inmediata, se calcula como los activos circulantes menos el inventario menos los pasivos circulantes.

Razón de apalancamiento. Miden el uso de la deuda por parte de la empresa, comparado con el uso de capital y su capacidad de pagar los intereses y otros gastos fijos.



Razón de cobertura del servicio de la deuda. La razón de deuda total toma en cuenta todas las deudas de todos los vencimientos de todos los acreedores.

Razón de liquidez. El punto de interés principal es la capacidad de la empresa para pagar sus cuentas en el corto plazo sin presiones excesivas.

Razón de rentabilidad. Tienen como finalidad medir la eficacia con que las empresas usan sus activos y la eficiencia con que administran sus operaciones.

Valor del mercado del patrimonio neto. El valor de mercado es el precio al que los compradores y los vendedores están dispuestos a negociar los activos [19].



# Bibliografía

- [1] Agresti Alan, *Categorical Data Analysis*, John Wiley & Sons, Inc, (2002).
- [2] Barboza César, Carbajal Braulio, *No tiene cuenta bancaria 56 % de la población en México*, Milenio, Negocios, 15 de junio de 2016. Consultado en: <http://www.milenio.com/negocios/>
- [3] Bonilla María, Olmeda Ignacio, Puertas Rosa, *Modelos Paramétricos y no paramétricos de Credit Scoring*, Revista Española de Financiación y Contabilidad, (2003).
- [4] Cuadras Carles M., *Nuevos Métodos de Análisis Multivariante*, CMC Editions, (2014).
- [5] Encino Víctor, *El sistema de las instituciones financieras no comerciales en Taiwán*, Taiwan Studies, Faculty Research Award Program for Paraguayans, (2005).
- [6] Escobedo Portillo M.T., Salas P. M. Jorge A., *P. Ch Mahalanobis y las aplicaciones de su distancia estadística*, CULCyT, (2008).
- [7] Falkenstein E., *RiskCalc For Private Companies: Moody's Default Model. Rating Methodology*, Moody's Investor Service, Global Credit Research, (2000).
- [8] González A. Roberto, *Se disparó en enero la cartera vencida de crédito al consumo*, La Jorna-

- da, Economía, 8 de marzo de 2017. Consultado en: <http://www.jornada.unam.mx/2017/03/08/economia/024n1eco>
- [9] Gutiérrez Girault Matías Alfredo, *Modelos de Credit Scoring - Qué, Cómo, Cuándo y Para Qué*, MPRA Paper, University Library of Munich,(2007).
- [10] Hand D. J., Henley W. E., *Statistical Classification Methods in Customer Credit Scoring: A Review*, Journal Royal Statistical Society A, (1997).
- [11] Haykin S., *Neural Networks: A Comprehensive Foundation*, Prentice Hall, (1994).
- [12] Hosmer D. & Lemeshow S., *Applied Logistic Regression*, John Wiley & Sons, (2000).
- [13] Johnson Richard A., Wichern Dean W., *Applied Multivariate Statistical Analysis*, Pearson, (2002).
- [14] Lichman M., *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml>], University of California, School of Information and Computer Science, (2013).
- [15] Mester J. Loretta, *What is the point of Credit Scoring.*, Business Review Federal Reserve Bank of Philadelphia, (1997).
- [16] Nasr G. E., Badr E. A., *Cross Entropy Error Function in Neural Networks: Forecasting Gasoline Demand*, Conference Paper, (2002).
- [17] Nieto Murillo Soraida, Pérez Salvador Blanca Rosa, Soriano Flores José Fernando, *Crédito al Consumo: La estadística aplicada a un problema de Riesgo Crediticio*, Revista Mexicana de Investigación Actuarial Aplicada, Colegio Nacional de Actuarios, (2010).
- [18] Pozo Díaz Maylí Z., Carrasco O. Gonzalo I., *Aplicación del Análisis Discriminante a un conjunto de datos vinícolas mediante el paquete estadístico SPSS v10*, Tecnociencia Vol. 7, (2005).

- 
- [19] Ross A. Stephen, Westerfield W. Randolph, Jaffe F. Jeffrey, *Finanzas Corporativas*, McGrawHill, Novena Edición, (2012).
- [20] Real Academia Española, *Diccionario de la lengua española (23.a ed.)*, Consultado en <http://www.rae.es/rae.html> (Octubre 2016).
- [21] SPSS (2016), *IBM SPSS Statistics 24 para Windows*.
- [22] Wang Eric, *Taiwan's Credit Card Crisis*, Recuperado de: <http://sevenpillarsinstitute.org>, (Noviembre 2016).
- [23] West David, *Neural Network Credit Scoring models*, Computer and operations research, in Computers & Operations Research, (2000).
- [24] Yeh I. C., Lien C. H., The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients., in Expert Systems with Applications, (2009).
- [25] Zhang Heping, Singer B.H, *Recursive Partitioning and Applications*, Springer, Second Edition, (2010).
- [26] Zorrilla Arena Santiago, *Cómo aprender Economía*, Limusa, Tercera Edición, (2007).