



# UNIVERSIDAD DE QUINTANA ROO

---

---

**División de Ciencias Sociales y Económico Administrativas**

UN MODELO DE *CREDIT SCORING* APLICADO A LOS  
MICROEMPRESARIOS MEXICANOS

**Tesis**

**Para obtener el grado de:**

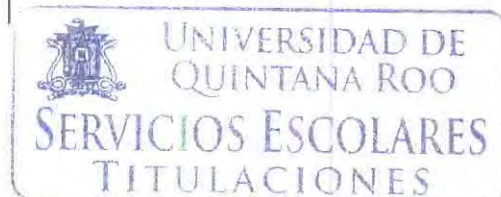
**LICENCIADA EN ECONOMÍA Y FINANZAS**

**Presenta**

**Tiaré Montserrat Manzanero Noh**

**Director de Tesis**

**Dra. Verónica Patricia Rodríguez Vázquez**



**Chetumal, Quintana Roo, México, Enero 2017.**





## UNIVERSIDAD DE QUINTANA ROO

### División de Ciencias Sociales y Económico Administrativas

“Un modelo de *Credit Scoring* aplicado a los microempresarios mexicanos”.

Presenta: **Tiaré Montserrat Manzanero Noh**

Tesis elaborada bajo la supervisión de comité de asesoría y aprobado como requisito parcial para obtener el grado de:

### LICENCIADA EN ECONOMÍA Y FINANZAS

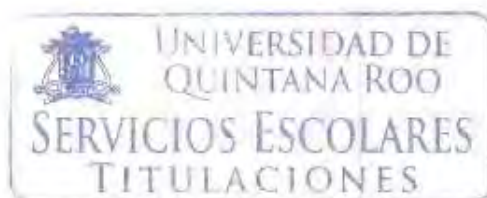
#### COMITÉ DE SUPERVISIÓN

Director: Dra. Verónica Patricia Rodríguez Vázquez

Asesor: Mtro. Naiber José Bardales Roura

Asesor: Dr. José Luis Esparza Aguilar

Chetumal, Quintana Roo México, Enero 2017.



## **Dedicatoria**

### ***A Dios***

Por haberme permitido llegar hasta aquí, estando conmigo en cada paso que doy y haber puesto en mi camino a aquellas personas que han sido soporte y compañía para esta etapa.

### ***A mis padres***

Por todo su cariño, por ser el pilar fundamental en todo lo que soy, por su incondicional apoyo a través del tiempo, por la motivación y enseñanzas de todos los días. Este logro es de ellos.

### ***A mi familia***

A mis hermanas Andrea y Elyssa, a mis abuelitas Cristina y Ligia, por el apoyo que siempre me han brindado.

### ***A mis maestros***

A la Dra. Verónica Rodríguez y a los maestros que me motivaron para la culminación de mis estudios y para la elaboración de esta tesis.

### ***A Juan Carlos***

Por su paciencia, apoyo y amor de todos los días, pero sobre todo por estar siempre a mi lado y ser alguien importante para lograr esto.

# Contenido

<b>I. Introducción .....</b>	<b>6</b>
I.1. Antecedentes del tema _____	6
I.2. Planteamiento de problema _____	9
I.2.1. Información asimétrica.....	9
I.3. Justificación _____	10
I.4. Objetivos de la investigación _____	11
I.4.1. Objetivo general.....	11
I.4.2. Objetivos específicos.....	11
I.5. Hipótesis _____	11
<b>II. Marco Teórico.....</b>	<b>12</b>
II.1. Microfinanzas _____	12
II.2. Riesgos financieros _____	13
II.3. Riesgo de crédito _____	13
II.4. Modelos de <i>Credit Scoring</i> _____	14
II.4.1. Análisis Discriminante.....	14
II.4.2. Modelos logísticos .....	15
II.4.3. Redes neuronales.....	17
II.4.4. Otros métodos .....	19
II.4.5. Comparación de algunos modelos. ....	20
<b>III. Metodología.....</b>	<b>22</b>
III.1. Modelo de Regresión logística _____	23
III.2. El programa <i>Neuraltools</i> de Palisade. _____	24
III.3. Construcción de la Base de Datos _____	26
<b>IV. Regresión Logit.....</b>	<b>33</b>
IV.1. Estimación del modelo Logit _____	33

IV.2. Interpretación de resultados	38
IV.2.1. Impactos	38
IV.2.2. <i>Odds ratio</i>	40
IV.2.3. Probabilidades con el modelo Logit	43
<b>V. <i>Neuraltools</i></b>	<b>46</b>
V.1. Aplicación de <i>Neuraltools</i>	46
V.2. Aplicación <i>Neuraltools</i> a los datos de los microempresarios.	47
<b>VI. <i>Discusión de resultados</i></b>	<b>57</b>
VI.1. Regresión logística	57
VI.2. <i>Neuraltools</i>	59
VI.3. Comparación de resultados	60
<b>VII. <i>Conclusión</i></b>	<b>64</b>
<i>Bibliografía</i>	<b>66</b>
<i>Anexos</i>	<b>71</b>

# I. Introducción

## I.1. Antecedentes del tema

Desde años atrás el riesgo es un tema importante para el sector financiero. Según la definición del Banco de México [BANXICO] (2005) el riesgo está relacionado con la posibilidad de que ocurra un evento que se traduzca en pérdidas para los participantes en los mercados financieros, como pueden ser inversionistas, deudores o entidades financieras. Existen diferentes tipos de riesgos financieros como el de liquidez, el de mercado, riesgo de tasas de interés, riesgo de cambio, pero el más importante para este trabajo son el riesgo de crédito y el de contraparte.

Este último riesgo ocurre cuando una de las partes de un contrato financiero, no cumple con las obligaciones acordadas, provocando que la parte contraria tenga una pérdida. En riesgo del crédito es cuando una persona que haya solicitado un préstamo no cumple con el pago de este.

Existen diversos factores para medir el riesgo del crédito:

- Probabilidades de incumplimiento.
- Las correlaciones entre incumplimientos.
- La exposición a cada deudor.
- La tasa de recuperación en caso de incumplimiento de los deudores.

La más importante para este trabajo es la probabilidad de incumplimiento y es la medida de que tan probable es que un acreditado deje de cumplir con sus obligaciones contractuales, tomando únicamente valores entre cero y uno. (BANXICO, 2005).

De igual manera es importante conocer el significado de un crédito, este hace referencia a un préstamo de una cantidad monetaria donde la persona que lo solicita se compromete a pagar en tiempo y forma, según con los lineamientos que ambas partes hayan acordado, el monto que le fue prestado más una tasa de interés y otros costos asociados. De manera general el crédito es un acuerdo

entre dos partes donde una recibe dinero, bienes o servicios y se compromete a pagarlos en un futuro, la otra parte recibe de nuevo lo que presto más una ganancia que es el interés.

Existen diferentes tipos de créditos:

- Comercial: préstamo otorgado a empresas para la adquisición de bienes, refinanciamiento de deuda, pago de servicios o algún otro gasto que tengan.
- Hipotecario de vivienda: está destinado a la adquisición de un terreno para la construcción de una vivienda, al igual que para su remodelación o mejoramiento.
- Crédito de consumo: está destinado a financiar como su nombre lo dice bienes de consumo o pago de servicios, donde la fuente principal de ingresos de este, son los sueldos de las personas.
- Microcrédito: son préstamos para actividades económicas de pequeña escala.

Debido a la existencia del crédito y el riesgo que se corre por otorgarlo y que este no sea devuelto, las entidades que se dedican al financiamiento optaron por utilizar métodos que disminuyeran este riesgo. Desde años atrás la banca ya empleaba algunos de estos métodos, pero hasta hace no muchos años empezó a emplearse en las instituciones de microfinanzas.

Tomando más auge con los acuerdos de Basilea, donde el último (Basilea III) surge como respuesta a la crisis financiera internacional, este acuerdo es un conjunto de reformas elaborado por el Comité de supervisión bancaria de Basilea ya que se tenía la necesidad de fortalecer la regulación, supervisión y gestión de riesgos del sector financiero. El Comité de Basilea acordó el marco de Basilea III en septiembre del 2009 persiguiendo las siguientes medidas:

- Mejorar la capacidad del sector bancario para afrontar perturbaciones ocasionadas por tensiones financieras o económicas de cualquier tipo.
- Mejorar la gestión de riesgos y el buen gobierno en los bancos.

- Reforzar la transparencia y la divulgación de información de los bancos.

Y las reformas se dirigen a:

- La regulación de los bancos a títulos individual (dimensión microprudencial), para aumentar la capacidad de reacción de cada institución en periodos de tensión
- Los riesgos sistémicos (dimensión macroprudencial) que puedan acumularse en el sector bancario en su conjunto, así como la amplificación procíclica de dichos riesgos a lo largo del tiempo.

Lo que esto ocasionó fue que las instituciones financieras buscaran formas de protegerse y reducir su riesgo de crédito a través de métodos calificadores comúnmente llamados *credit scoring*.

El *credit scoring* es un sistema relativamente nuevo que llegó para mejorar la eficiencia de los informes hechos a mano, que contaban con muchos errores, obteniendo resultados neutros y sólidos hablando estadísticamente. El *credit scoring* son todas las técnicas y modelos estadísticos que ayudan a los prestamistas para el otorgamiento de crédito principalmente de consumo, midiendo el riesgo de prestarle a un determinado cliente. (Dabós, 2010) Otra definición es de Hand y Henley (1997) quienes mencionan que son procedimientos estadísticos que se usan para clasificar a aquellos que solicitan crédito, inclusive a los que ya son clientes de la entidad crediticia, en los tipos de riesgo bueno y malo.

Al clasificar a los clientes entre buenos y malos, lo que hacen las entidades financieras es reducir su riesgo de pérdidas por incumplimientos, porque además de darles un puntaje a los solicitantes de un crédito, ayuda a saber que a tasa de interés realizar el préstamo, el *credit scoring* permite a las entidades financieras tomar decisiones rápidas y objetivas.



## I.2. Planteamiento de problema

Debido al riesgo que corren las instituciones financieras al otorgar préstamos en México, lo que se pretende realizar es un modelo de *credit scoring* para conocer las características específicas con las que cuenta un microempresario mexicano y así asignarles una calificación y las probabilidades de incumplimiento de pago, la intención es que este modelo sirva como una herramienta para las instituciones financieras y así sea más fácil tomar una decisión al momento de otorgar créditos ya que se contaría con calificaciones para cada tipo de cliente y podría evitarse y minimizar los clientes que tiene un alto riesgo de incumplimiento y así la empresa no obtenga pérdidas.

### I.2.1. Información asimétrica

Esta situación hace referencia a las situaciones en la que la información no es igual para ambas partes que conforman una transacción, es decir que uno de los agentes involucrados tiene más información que otro. Por lo tanto es de esperarse que los que cuenten con más información son los que tomaran mejores decisiones y los menos informados seguramente son los que sufrirán pérdidas. La asimetría lleva a dos problemas

- La selección adversa: surge cuando una parte de la transacción posee información relevante desconocida por la otra, de forma que la primera realiza el proceso decisional y la transacción en superioridad de condiciones. la parte menos informada no es capaz de distinguir la buena o mala "calidad" de lo ofrecido por la otra parte, por lo que acaba suponiendo que lo más probable es la posibilidad peor, ya que, por lo general, es lo más interesante para el otro agente. (Gómez, 2008)
- Riesgo moral: Se considera que existe un problema de riesgo moral (o acción oculta) cuando una parte de la transacción puede adoptar determinadas acciones que afectan a la valoración que la otra parte ha hecho de la transacción, pero que esta no puede controlar perfectamente.

(Gómez, 2008) Este problema de riesgo moral generalmente se da de manera ex-post sobre las actuaciones futuras de los involucrados.

### I.3. Justificación

Desde años atrás, se ha llevado a cabo la medición del riesgo para para la banca, pero en los últimos años se ha implementado más en las instituciones que otorgan financiamiento, ya que es importante conocer para ellas las características con la que contaría un buen cliente y así reducir los riesgos que corren al otorgar préstamos a personas que no cumplen con el pago de estos.

Por lo anterior es importante que las instituciones cuenten con un método que sirva como herramienta de apoyo al momento de tomar decisiones sobre el otorgamiento de créditos y así evitar pérdidas en las empresas.

Los modelos de *credit scoring* pueden impulsar el crecimiento económico, aumentar el acceso a recursos esenciales por parte de los consumidores y permitir una asignación más eficiente del riesgo (Pagano y Jappelli, 1993) además de que son esenciales para resolver tres problemas económicos: (Turner y Robin, 2006)

- Niveles de eficiencia inferiores a estándares internacionales en el sector financiero.
- El relativo estancamiento de los préstamos del sector privado.
- El riesgo de crisis financieras, que a menudo derivan, en parte, de los problemas de selección adversa en el sector bancario

La metodología seleccionada para realizar el modelo de *credit scoring* son las redes neuronales artificiales, ya que según Pérez y Fernández (2007) son superiores a otras técnicas porque entrenan, auto organizan, aprenden y olvidan, son robustas y tolerantes a las fallas, son flexibles pudiendo adaptarse a diferentes ambientes, además de ser hábiles en el proceso de asociar, evaluar o reconocer patrones.

## I.4. Objetivos de la investigación

### I.4.1. Objetivo general

Realizar un modelo de *credit scoring* a través del método de Redes Neuronales Artificiales (RNA) con el cual puedan conocerse las características que influyen en los microempresarios mexicanos para que cumplan con el pago de créditos, todo esto en base a datos generados de la ENAMIN (Encuesta Nacional de Micronegocios) 2012.

### I.4.2. Objetivos específicos

- Conocer las características específicas de los microempresarios mexicanos que cumplen con los pagos de un crédito que se les fue otorgado.
- Conocer la probabilidad de incumplimiento de un microempresario dada ciertas características.

## I.5. Hipótesis

La construcción del modelo de *Credit Scoring* para microempresas mexicanas, permitirá detectar los factores de cumplimiento o incumplimiento en el pago de los créditos, por lo que se podrá constatar la siguiente hipótesis:

H<sub>0</sub>: Los factores de cumplimiento del pago de un crédito por parte de las microempresas mexicanas en 2012, involucran factores como: sexo, tipo de crédito, edad y clasificación por pobreza de la zona geográfica.

## II. Marco Teórico

### II.1. Microfinanzas

Las microfinanzas son una herramienta utilizada para tratar de disminuir la pobreza ya que el acceso a estos servicios financieros, hace que personas de escasos recursos puedan aumentar sus ingresos, teniendo un mejor manejo sobre sus finanzas. Según el *Consultative Group to Assist the Poor* (CGAP), las microfinanzas se refieren a sistemas financieros que atienden las necesidades de las personas. Existen instituciones microfinancieras que otorgan créditos, ahorros, seguros y transferencias de dinero a microempresarios para así darles un apoyo y ellos puedan sostener sus actividades productivas.

Lo más conocido de las microfinanzas es el microcrédito, el cual según la Comisión Nacional para la Protección y Defensa de los Usuarios de Servicios Financieros [CONDUSEF] (2012) es un préstamo dirigido fundamentalmente a personas que quieren montar un pequeño negocio o autoemplearse para poder desarrollar alguna actividad productiva independiente. Existen dos tipos de microcrédito, el individual y el grupal.

- Individual: en un esquema similar a los créditos de la banca tradicional. Es un financiamiento que se otorga a una persona para actividades productivas, la atención por parte de la institución es muy personalizada y de acompañamiento a los clientes. Para este crédito es necesario contar con un aval o algún tipo de garantía prendaria. Es muy común que los montos de los préstamos individuales sean más altos que el de los grupales. (CONDUSE, 2012)
- Grupales: Están dirigidos a personas que al no contar con garantías propias pueden formar grupos de más de cinco personas de la misma comunidad para responder en conjunto por el crédito. Al formar estos grupos, las personas deben de conocerse ya que se otorga un crédito a cada persona pero si una de ellas no cumple con el pago, el grupo se hace responsable por el pago. (CONDUSEF, 2012)

## II.2. Riesgos financieros

Los riesgos financieros están relacionados con las pérdidas en el valor de un activo financiero, tales como un préstamo o una inversión (Martínez, 2012), en general está asociado a cualquier forma de financiación. Se puede entender al riesgo como la posibilidad de los beneficios o ganancias que se obtienen sean menores a los esperados o no se obtenga nada. Existe una clasificación para estos riesgos y alguno de ellos son los siguientes:

- Riesgo de crédito: se da cuando una de las partes de un contrato financiero no asume las obligaciones pactadas sobre el pago.
- Riesgo de liquidez: se produce a consecuencia de continuas pérdidas en la cartera. Una de las partes tiene activos pero no tiene la liquidez suficiente para atender sus obligaciones. Existen dos tipos de riesgo de liquidez, la liquidez de activos y la de financiación.
- Riesgo de mercado: Son las operaciones enmarcadas en los mercados financieros, se da cuando el valor de una cartera de inversión o negociación, se reduce debido al cambio en los valores de los factores de riesgo del mercado, los cuales son Riesgo de tipo de interés, riesgo cambiario, riesgo de mercancía y riesgo de mercado.
- Riesgo operacional: es el riesgo relacionado con las actividades en las que incurre la empresa o comercio.

## II.3. Riesgo de crédito

El riesgo de crédito surge cuando las contrapartes están indispuestas o son totalmente incapaces de cumplir sus obligaciones contractuales. El riesgo de crédito abarca tanto el riesgo de incumplimiento, que es la evaluación objetiva de la probabilidad de que una contraparte incumpla, como el riesgo de mercado que mide la pérdida financiera que experimentada si el cliente incumple (Tellez, 1999).

Clasificación del riesgo de crédito:

- Soportado por personas físicas: Las personas o individuos están expuestas a este riesgo, ya que puede suceder que la empresa o institución no cumpla con las obligaciones pactadas.
- Soportado por las empresas: cuando un cliente hace una compra a plazos de una mercancía, la empresa corre el riesgo de la persona no pague es por eso que las empresas cuentan con departamentos de valoración de riesgo.
- Soportado por las instituciones financieras: lo más conocido es la concesión de crédito a clientes, ya sea individuales o corporativos. Al otorgar estos créditos las instituciones corren el riesgo de que el solicitante del préstamo no cumpla con el pago de su deuda.

El riesgo de crédito cobra gran importancia debido al constante crecimiento de la globalización y los cambios que se generan en el mercado financiero, lo da lugar a que cada día se intente mejorar los mecanismos o métodos que buscan reducir en gran medida todo tipo de riesgos financieros.

Una técnica muy utilizada en los últimos años, para reducir el riesgo crediticio que incurren las empresas es a través de los modelos de *credit scoring* los cuales otorgan una calificación para cada tipo de cliente, logrando que las instituciones financieras puedan tomar decisiones en relación a esta calificación y así disminuir en mayor medida el riesgo.

## II.4. Modelos de *Credit Scoring*

Los modelos de *credit scoring* se pueden elaborar con distintas herramientas paramétricas tales como los modelos logit y probit, el análisis discriminante y no paramétricas como árbol de toma de decisión, máquina de vectores, juicio humano y redes neuronales entre otras.

### II.4.1. Análisis Discriminante

El primer modelo de *credit scoring* que se desarrolló para las microfinanzas fue realizado por Vigano (1993), para una institución de microfinanzas de Burkina

Faso, utilizando el método de análisis discriminante para elaborar el modelo. Este método tiene como objetivo analizar la relación entre una variable dependiente cualitativa y un conjunto de variables independientes cuantitativas, a partir de una serie de funciones discriminantes, aunque tiene la incapacidad para calcular las probabilidades de incumplimiento de pago (Escalona, 2011). De igual manera autores como Ocha, Galeano y Agudelo (2010) desarrollaron un modelo con el método de análisis discriminante con datos proporcionados por una cooperativa financiera del Valle de Aburra, para definir perfiles de prestatarios propensos al incumplimiento del pago y prestatarios que si cumplen con el pago.

Ortega, Martínez y Valencia (2009) presentan un método denominado Z-Score de Edward I. Altman, donde se utiliza la técnica de análisis discriminante tomando como datos a los indicadores financieros de las compañías HB Fuller Colombia de los años entre el 2001 y el 2005. Este procedimiento consistía en identificar combinaciones lineales de los indicadores o ratios que son más significativos de las compañías, para así poder clasificar a los clientes entre buenos pagadores o no. Calin y Popovici (2014) de igual manera desarrollan el Z-Score de Altman, explicado lo que el desarrollo a diferencia de los otros autores que lo llevaron a la práctica, los dos autores anteriores mencionan que este modelo tiene una serie de deficiencias que no permite dar buenas predicciones, se construye de manera lineal, y en ocasiones sería mejor realizarlo de una manera no lineal.

#### II.4.2. Modelos logísticos

Otro de los métodos es la regresión logística (logit) que se utiliza cuando lo que se quiere pronosticar es la probabilidad de ocurrencia de un suceso determinado, en esta a diferencia del análisis discriminante se pueden utilizar variables independientes cuantitativas y se puede medir el impacto de cada una de ellas con la variable dependiente, además que es un método muy sencillo de interpretar e implementar. Autores como Coloma y Weber (2006) desarrollaron modelos logit, en este caso para el Instituto de Desarrollo agropecuario del Chile (INDAP), el cual es el encargado de facilitar el acceso al financiamiento a los agricultores de este

país. El problema fundamental que quisieron resolver en su trabajo fue ¿Cuál es la probabilidad de que un crédito no se recupere? en el INDAP y lo resolvieron en dos etapas, la primera era la segmentación del universo de créditos /clientes y la segunda la construcción de modelos predictivos logit para cada segmento.

En Schreiner (1999) realiza un modelo logit que fue aplicado a una organización de microfinanzas en Bolivia, dándole como resultado que el modelo que realizó tiene menos poder predictivo que la mayoría de los modelos de calificación que se utilizan en compañías que otorgan créditos además que se le puede dar una mejor uso a como un filtro que resalta lo casos de clientes que tiene que revisarse cuidadosamente. Gutiérrez (2007) realiza un trabajo donde menciona que los modelos logit son superiores a los modelos lineales ya que dan para cada deudor un probabilidad “default”. También menciona que los modelos de *credit scoring* se emplean mayormente para evaluar individuos y pequeñas y medianas empresas. Las grandes empresas utilizan un sistema llamado rating.

Lara, Rayo y Camino (2010). Realizan un modelo de regresión logística binaria y para la aplicación de este *scoring* utilizaron dos métodos mencionados en el acuerdo de Basilea II. El primero es un enfoque estándar donde se asigna una ponderación de riesgos a cada uno de los activos y el segundo está basado en ratings internos (modelo IRB) basándose en las estimaciones propias de la entidad financiera, considerando dos niveles: método basado en calificaciones internar IRB básico y avanzado.

Torrico (2014) propone la aplicación de un modelo de *credit scoring* para medir el riesgo crediticio agregado de todos los bancos de Bolivia, trabajando a nivel macro ya que está hablando de un agregado y no como los otros que se realizan a nivel micro (individuos), todo esto a través de los modelos logit.

Bravo, Maldonado y Weber (2010) de igual manera desarrollan un modelo de regresión logística para definir si un cliente que cuenta con ciertas características será capaz de cumplir con el pago de un crédito, pero a través de una metodología



utilizada para los datos llamada KDD (*Knowledge Discovery in Databases*), el cual es un proceso completo de extracción del conocimiento de base de datos, al utilizar este proceso se tiene menos errores y por lo tanto un mejor modelo.

Espin y Rodriguez (2013) desarrollaran una metodología para elaborar un modelo que logre predecir el comportamiento de impago en función de la información sociodemográfica de los clientes. A través de un híbrido de dos técnicas, ya que se combinan las técnicas de análisis multivariado (segmentación: árbol de decisión) con una regresión logit.

### II.4.3. Redes neuronales

Lo que realizan estas redes es imitar el funcionamiento del cerebro humano, logrando construir sistemas con alto grado de inteligencia. Juan Lazo y Cristian López realizan un modelo para la detención de riesgo crediticio, donde se utilizaron un total de 19 modelos neuronales. Porto y Castromán (2000) en sus investigaciones mencionan que las redes neuronales lo que realizan es suplantar datos que hicieran falta, ya que reproducen el funcionamiento de la técnica que pretende complementar y es una buena metodología para aplicar a una política de una empresa.

Serrano y Gallizo (1996) mencionan que las redes neuronales tratan de resolver de forma eficiente los problemas que pueden situarse en tres grupos: optimización, reconocimiento y generalización. Y esto es muy aplicable al campo de la gestión empresarial.

Por otro lado Pérez y Fernández (2007) presentan una lista de ventajas y desventajas de utilizar esta metodología:

Ventajas:

- Se entrenan, auto organizan, aprenden y olvidan.
- Son robustas y tolerantes a fallas; la falla de una o varias neuronas no implica un fallo total en la red neuronal

- Son flexibles, lo que les permite adaptarse fácilmente a nuevos ambientes, ya que pueden catalogarse como sistemas independientes.
- Se emplean en datos en los cuales el patrón es oscuro e imperceptible, que exhiben comportamiento impredecible o no lineal, como en modelos tradicionales de series temporales y datos caóticos.
- La velocidad de respuesta es menor que la del cerebro humano.
- Son hábiles en el proceso de asociar, evaluar o reconocer patrones

#### Desventajas

- Funcionan como una caja negra, resuelven un problema, pero es difícil saber cómo lo han hecho.
- No resuelven todos los problemas ni siempre los resuelven de la mejor manera.
- Tienen problemas en la estimación de cálculos precisos. Funcionan bien con problemas complejos de difícil cálculo pero que no requieren respuestas perfectas, sino solo respuestas rápidas y buenas.
- En predicción bursátil se emplean con frecuencia en boletines divulgativos de las entidades financieras, lo que significa que su éxito es relativo, puesto que en otro caso su predicción no se publicaría, sino que se aprovecharía.
- Las redes neuronales se están comportando bien en predicción a largo plazo con componentes no lineales; en cambio, no están claras las mejorías observadas en series cortas y estacionales, como son típicas en las predicciones de ventas.

Otro autores que han trabajado con redes neuronales aplicado a calificación crediticia son Ariza, Barón, Obregón, Pineda y Velosa (2011) realizaron un trabajo que tenía como, identificar las variables más relevantes que permiten a través de un modelo de redes neuronales artificiales (RNA), alcanzar los mejores predictores de bajo nivel de riesgo, en la etapa de otorgamiento de microcréditos, esto aplicado en el país de Colombia.

#### II.4.4. Otros métodos

Existen otros métodos conocidos como el árbol de toma de decisión, Cardona (2004) menciona que el uso de esta herramienta efectivamente sirve la para la predicción de probabilidades de incumplimiento, no solo a nivel de capacidad de discriminación, si no como una herramienta que es muy fácil comprender y puede utilizarse para realizar planeación de estrategias comerciales, ya sea de ventas, servicios, estrategia de cobranzas entre otras.

Otra metodología es el de máquinas de vectores de soporte, Moreno y Melo (2011) quienes han utilizado este método afirman que requiere menos supuestos sobre los datos de entrada en comparación con otros modelos, su desempeño no depende del tamaño o dimensión de la muestra, se resuelve mediante programación cuadrática, el algoritmo se ajusta a problemas no lineales, además que el SVM tiene una buena capacidad de pronóstico.

Otros autores como Beltrán, Muñoz Martinez y Muñoz Alamillos (2014) han utilizado redes bayesianas para realizar modelos de *credit scoring*. Lo que se hace en este enfoque es que de cada situación que se presenta siempre se va a elegir la acción que maximiza la utilidad esperada. Estos modelos bayesianos pueden utilizarse para resolver problemas con una perspectiva descriptiva así como una predictiva, en la descriptiva se basan en descubrir las relaciones de dependencia/independencia. En la predictiva se toman a las redes bayesianos como métodos de clasificación.

Gómez (2012) utiliza la lógica difusa para desarrollar un modelo de calificación de crédito. La finalidad de esta metodología es tener una lógica que se aproxime al pensamiento real del ser humano, con el uso de un lenguaje natural y unas variables amplias y vagas para ser consideradas naturales. Este método puede funcionar como una herramienta complementaria para los modelos que ya existen para obtener más detalles de la calificación crediticia.

Otro enfoque como herramienta de *credit scoring* que mide las probabilidades de incumplimiento y las tasas de recuperación, es el enfoque de Merton y sus

extensiones realizado por primera vez en 1974. Cerezo, Claramuunt y Casanovas (2011) realizan la aplicación de este modelo a la bolsa de valores de Colombia.

Banda y Garza (2014) presentan un modelo llamado Holt-Winters, el cual es utilizado mayormente para predecir y generar intervalos de confianza sobre posibles ingresos y para pronosticar series de tiempo que tengan las siguientes características: 1) Nivel, 2) Tendencia aditiva y 3) estacionalidad multiplicativa.

#### II.4.5. Comparación de algunos modelos.

Existen distintos autores que han realizado varios modelos para comparar su poder predictivo sobre el riesgo de crédito de diversas instituciones. Unas de ellos son Martí y Puertas (2012) realizaron un análisis de la capacidad predictiva de tres modelos, dos paramétricos y uno no paramétrico. Para este análisis empírico aplicaron para los modelos paramétricos el AD (análisis discriminante) y el Logit y para el no paramétrico el algoritmo CART (árbol de decisión). En su trabajo se menciona que el modelo AD es una técnica estadística multivariante que permite estudiar de forma simultánea el comportamiento de un conjunto de variables independientes, para así identificar segmentos o grupos en una serie de datos previamente determinados y excluyentes. El modelo logit permitió calcular la probabilidad de que un individuo pertenezca a o no a uno de los grupos establecidos. Y por último el modelo no paramétrico CART que son árboles de decisión, sirve para separar las observaciones de una muestra asignándole a cada una un grupo, de esta manera se minimiza el coste esperado de los errores. Además llega a la conclusión que los modelos no paramétricos son más flexibles que los modelos paramétricos ya que no necesitan una forma funcional, ya que se ajustan a la que mejor se aproxime a las variables del modelo.

Ladino (2014) realizó una comparación del poder predictivo de las regresiones logit y redes neuronales utilizando funciones de costos, aplicado a una institución financiera colombiana. Para él es mejor elaborar un modelo de redes neuronales ya que los estadísticos con los que cuenta en el modelo realizado son mayores

que los de la regresión logística, pero es importante tomar en cuenta que es más complicado elaborar modelos de redes neuronales. Las redes neuronales son mejores ya que tiene un mayor poder de discriminación y un impacto económico material.

Así como existen autores que comparan los diversos modelos posibles para elaborar un *credit scoring*, hay autores como Esteve (2007) que juntan dos modelos para obtener un mejor resultado, en su trabajo realizado esta autora cálculos de análisis discriminante, el programa Matlab para aplicar el algoritmo de Kohonen al igual que el uso de *toolbox Neural Network*. Comprobando que los resultados del *credit scoring* con la conjunción de modelos estadísticos tradicionales y redes neuronales son mejores que los que se obtienen por si solos en cada modelo.

Después de revisar los diferentes modelos de *credit scoring* que anteriormente fueron mencionados, junto con sus autores y su forma de trabajo; considerando que estos fueron desarrollados en diferentes países y presentan mucha diversidad de información se toman como base y apoyo para poder elaborar dos modelos de *credit scoring*, uno con la regresión logística y otro utilizando redes neuronales. Los cuáles serán presentados y analizados en los siguientes apartados de este trabajo. Es importante mencionar que aunque los modelos analizados no han trabajado con datos de microempresarios, la metodología para elaborarlos sigue siendo muy similar.

### III. Metodología

Ya que se pretende analizar el comportamiento de pago de un cliente de micro crédito, se requiere contar con una base de datos amplia que cuente con el historial de préstamos que no han sido pagados. Para este trabajo se utilizarán datos de microempresarios mexicanos del año 2012, de acuerdo a la ENAMIN realizada por el INEGI en este mismo año. Todo esto con el fin de conocer que características influyeron en los microempresarios encuestados para que realicen o no el pago de su préstamo, además de obtener la probabilidad de incumplimiento dadas ciertas características. Así en un futuro suponiendo que los nuevos solicitantes de crédito tengan ciertas características iguales a los encuestados, se podrá reducir el riesgo de otorgar un crédito a una persona que podría incumplir con el pago. Para esta investigación el método con el que se elaborará el *credit scoring* será a través de las redes neuronales.

Las redes neuronales son un modelo de inteligencia artificial, donde lo que se pretende es que estas redes simulen la estructura y el comportamiento que tiene el cerebro humano. Utilizan procesos de aprendizaje para buscar una solución a diferentes problemas además de ser un conjunto de algoritmos matemáticos que encuentran las relaciones no lineales entre conjuntos de datos; suelen ser utilizadas como herramientas para la predicción de tendencias y como clasificadoras de conjuntos de datos. (Pérez y Fernández, 2007)

Las redes están formadas por una serie de procesadores a los cuales se les llama nodos y están todos interconectados, el objetivo de estos es mandar una señal que se le llama señal de entrada y obtener una señal de salida. Lara, Rayo y Camino (2010) mencionan que en el enfoque de *credit scoring* lo que realiza las redes neuronales consiste en considerar como nodos de entrada las características o variables de la operación d crédito, y de nodo de salida la respuesta definida como la probabilidad de ser un préstamo malo.

### III.1. Modelo de Regresión logística

De acuerdo con (Rojo 2007), el modelo de regresión logística parte de la hipótesis de que los datos siguen el siguiente modelo

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_k * x_k + u = x * b + u +$$

Para simplificar:

$$z = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_k * x_k$$

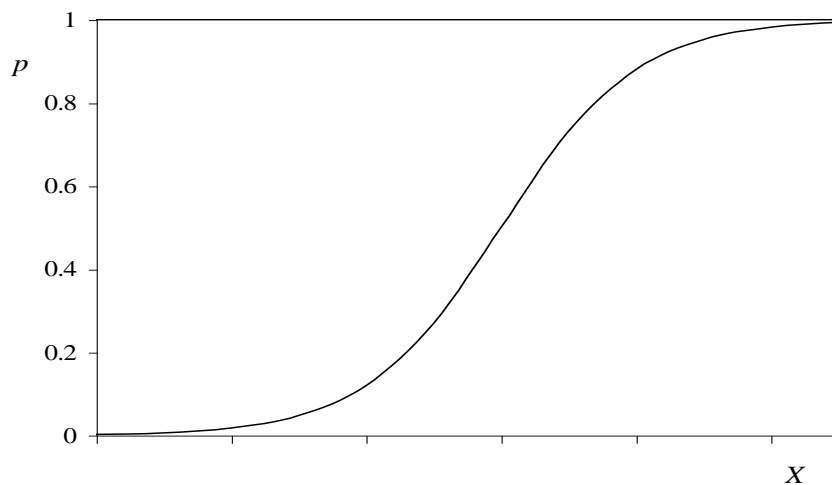
Por lo que el modelo se representa de la siguiente manera:

$$\ln\left(\frac{p}{1-p}\right) = z + u$$

Donde  $p$  es la probabilidad del suceso y es igual a:

$$p = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

Como se había mencionado antes, la regresión logit, no es una regresión lineal por lo que la distribución está representada con la siguiente imagen



Existen dos maneras de expresar el modelo de regresión logística, la primera es la logit y la segunda es llamada *odds ratio*, con estas formas es con la que se interpreta los coeficientes que se obtienen de resultado.

### III.2. El programa *Neuraltools* de Palisade.

La herramienta que se utilizó para poder desarrollar este trabajo fue un programa del grupo Palisade, el cual es un fabricante de software líder a nivel mundial en cuanto a cuestiones de análisis de riesgo y toma de decisiones, el nombre del programa es *Neuraltools*, el cual es muy sencillo de utilizar y es un auxiliar de redes neuronales que se utiliza para Microsoft Excel, con el cual se analizan datos que se presentan en las hojas de cálculo. Como se sabe las redes neuronales imitan funciones del cerebro, por lo cual pueden aprender complejas y diversas relaciones de datos, para luego dar predicciones para los nuevos datos.

Este programa es utilizado para resolver dos tipos de problemas, los de clasificación y los problemas numéricos, teniendo una gran gama de aplicaciones. Como por ejemplo en el mercado de valores, detecciones de fraudes crediticios, ventas de las empresas, riesgo en inversiones, diagnósticos médicos, investigación científica entre muchos más, pero el tema que interesa, que es para la asignación de créditos o préstamos. Este programa puede ser utilizado para resolver cualquier problema, siempre y cuando se tengan datos históricos y se quiera obtener una predicción, por lo cual *Neuraltools* cuenta con los modernos algoritmos de redes neuronales para así tener las mejores predicciones posibles, ya se aplicado a los problemas categóricos como es el caso de la asignación de créditos o préstamos y los problemas numéricos como pueden ser las predicciones en el mercado de valores.

*Neuraltools* trabaja a través de tres procedimientos básicos que son el entrenamiento, la prueba de las redes neuronales y la predicción utilizando las redes neuronales que se entrenaron. Para poder realizar los tres puntos anteriores es importante crear un conjunto de datos con las variables que se quieren trabajar, incluyendo los datos que se quieren predecir, es decir se creara un conjunto de



datos, con los datos históricos que se tengan y los nuevos datos que se quiera conocer su resultado. Para esto es importante tener identificadas las variables a utilizar, ya que el programa te permite clasificar en variables independientes ya sean numéricas o de clasificación y la variable dependiente de igual manera numérica o de clasificación.

Ya que se tenga creado el conjunto de datos y con las variables de acuerdo a su clasificación, se elige la red con la que se quiere trabajar, el programa de *Neuraltools* cuenta con 3 opciones de red, la primera es la Red PN/GRN (redes neuronales probabilísticas / redes neuronales de regresión generalizada), con esta selección si la variable dependiente es categórica se trabajara con la red PN y si es numérica con la GRN. Ambas redes trabajan de manera similar, ya que cada caso que entrenan lo representan con un nodo (elemento de la red). De esta manera la predicción de los valores desconocidos de la variable dependiente, se obtendrán mediante la interpolación de los casos que fueron entrenados. La segunda opción de red es la MLF (redes multinivel siempre hacia adelante), la cual consiste en ser una red de avance de múltiples capas (múltiples nodos), es decir que tiene una capa de nodos de entrada, uno o dos nodos en capas ocultas y una capa de salida, por lo tanto se cuenta con una configuración para determinar los nodos que se requieran. Y por último la tercera opción es búsqueda de la mejor red, con eso *Neuraltools* selecciona la red con la que mejor convenga trabajar.

Como se había mencionado antes, *Neuraltools* realiza tres procedimientos, el primero es entrenar los datos históricos con los que se cuenta, a través de la red que se haya seleccionado. Lo que realiza el programa en este procedimiento es generar una red neuronal con los datos conocidos que se tienen de las variables. Después de realizar el entrenamiento el programa prueba la red neuronal que se creó, para conocer qué tan acertadas son las predicciones, este entrenamiento se realiza con los datos históricos, aunque no se lleva a cabo con todos, mayormente solo se utiliza el 20% de estos. Al realizar la prueba, el programa arroja el porcentaje de respuesta que se obtuvo, ya sea correcta o incorrecta la predicción y elige la red que mejor porcentaje de respuesta correcta obtiene, para después con

esta llevar a cabo la predicción (con la red neuronal entrenada) para los datos que son nuevos y no se conoce sus valores en la variable dependiente, de igual manera como en la prueba, en este procedimiento se da el porcentaje de acierto correcto e incorrecto de la predicción. Mientras más datos son entrenados y probados, el programa necesita más tiempo, ya que se realiza el entrenamiento varias veces con diversas configuraciones para tener la red neuronal que ofrezca mejores resultados en las pruebas y así poder aplicarla a la predicción.

### III.3. Construcción de la Base de Datos

Los datos que se utilizaron para realizar este trabajo fueron tomados del INEGI con base en la “Encuesta Nacional de Micronegocios 2012”, la cual se conforma por cinco secciones, debido a que es una encuesta extensa solo se seleccionó lo más importante de cada sección, utilizando de esta manera las variables que podrían ser de utilidad. Como se había mencionado antes para este trabajo se decidió utilizar el programa de Palisade, Neural Tools. Por lo tanto para saber que variables era más convenientes utilizar, se llevó a cabo una regresión logística para conocer las variables que son significativas respecto a la variable dependiente.

Para llegar a las variables que se utilizaron, los datos de esta encuesta pasaron por una serie de filtros, el primero y quizá uno de los más importantes es que la información que se desea utilizar sea verdadera, ya que existen algunos casos donde se menciona que la información es falsa, debido a que el entrevistado no respondió. Como es un trabajo enfocado en los microempresarios que obtuvieron un préstamo, se realizó un filtro para trabajar solo con las observaciones donde las personas contaban con un crédito.

El tercer filtro fue seleccionar a los micronegocios que solo contaban con un dueño, lo anterior se hizo para facilitar la condición que se agrega sobre las ventas (se detalla más adelante). Por último se filtró a los negocios donde se conocía la mensualidad que pagaban por el préstamo, dejando fuera a los que no contaban

con esta información. En algunas observaciones no fue proporcionado este dato (mensualidad), pero con otros datos proporcionados se pudo calcular el pago que se realiza por el crédito mensualmente, utilizando una tasa del 3%.

Por otro lado se seleccionaron 26 variables de los 100 reactivos de la ENAMIN para trabajar con una regresión que nos ayudara a detectar la significancia estadística. Las primeras variables de clasificación son consideradas básicas e importantes debido a que representa la información general sobre las personas, las cuales son: la edad, el estado civil, el sexo y el nivel de estudios. Algunas de ellas se presentan de manera dicotómica, es decir, con dos valores (1 y 0).

Por ejemplo, el estado civil se dividió en dos, los que tiene pareja (1), ya sea que vivan en unión libre o casados y los que no tienen pareja (0), es decir los solteros, separados, divorciados y viudos.

Se aplicó lo mismo con el nivel de estudios, en este caso se dividió entre los que poseen una licenciatura o más (1) y los que no alcanzaron este nivel de estudios (0). En el caso del sexo solo pueden existir dos valores, hombres (1) y mujeres (0). Por último la edad se mantuvo como una variable cuantitativa medida en años. Otra variable básica es el estado donde se encuentra el microempresario, trabajar con cada uno de los estados, sería un poco complicado, por lo tanto se decidió dividir al país en 4 zonas, tal como el CONEVAL tiene hecho la división de acuerdo al porcentaje de población que sufren de pobreza en los estados. Del 20% al 35% se encuentran los estados de Baja California Sur, Baja California Norte, Nuevo León, Sonora, Chihuahua, Coahuila, Aguas calientes, Colima, Querétaro y Distrito Federal. Del 35% al 50% está conformado por los estados de Tamaulipas, Sinaloa, Durango, Nayarit Jalisco, Guanajuato, San Luis Potosí, Edo de México, Tabasco, Yucatán, Campeche y Quintana Roo. Para el rango de 50% a 65% se encuentra Veracruz, Morelos, Puebla, Hidalgo, Zacatecas, Michoacán, y Tlaxcala. Por último Oaxaca, Chiapas y Guerrero conforman el rango de 65 a 80%. En cada una de estas variables, el 1 representa cuando el rango es el mismo que el nombre de la variable y el 0 cuando es falso.

Las variables restantes se encuentran más relacionadas con el tema que se está trabajando, las cuales son consideradas importantes ya que pueden influir mucho al momento de poder pagar un préstamo. Como en el caso anterior, varias de ellas son variables dicotómicas. Seis de estas se derivan del tipo de institución financiera que otorgó el préstamo a los micronegocios, estas observaciones provienen de una sola pregunta que se realizó en la encuesta. Como la pregunta tiene ocho posibles respuestas y se busca saber qué tan significativas es cada una de ellas, lo que se hizo fue elaborar 6 variables dicotómicas, las cuales son las siguientes: banca comercial, banca de desarrollo, programa de gobierno, cajas populares, crédito de proveedores y préstamos con intereses.

En cada una de estas variables el 1 representa una respuesta verdadera (lo que indica que la persona recibió el préstamo de esta institución financiera a la que hace referencia) y el 0 una falsa (la persona recibió el préstamo de otra institución). Como se mencionó anteriormente, eran ocho respuestas, pero dos no fueron tomadas en cuenta ya que se alegó que no serían tan significativas, las cuales son préstamos de amigos o parientes y otros préstamos. Este grupo de variables son importantes porque se conoce si influye de dónde proviene el crédito para que las personas paguen o no. Por ejemplo, se puede pensar que un préstamo otorgado por un banco comercial tiene intereses muy altos en comparación con un crédito que otorga el gobierno, por lo que se podría decir que es más probable que paguen las personas que obtuvieron un préstamo por parte del gobierno.

Adicionalmente, tres variables dicotómicas son en referencia al giro del negocio debido a que es importante conocer si la actividad que se desarrolla es una característica para que el negocio pueda pagar el préstamo adquirido. La pregunta donde se tomaron las observaciones contaba con cuatro posibles respuestas que son: manufactura, comercio, construcción y servicios. Se decidió tomar tres respuestas como variables: manufactura, comercio y servicios, (donde el número 1 es verdadero y el 0 falso), quedando construcción cuando todas las variables

anteriores tengan 0. Las últimas variables dicotómicas fueron con relación al tipo de préstamo que habían adquirido las personas con un micronegocio.

Se crearon dos variables que son préstamos de habilitación o avío y préstamos perfeccionarios. Estas dos variables tienen el número 1 como respuesta verdadera y 0 como falsa. Cuando se tiene falso en las dos variables (0,0) hace referencia a que adquirieron otro tipo de préstamo o no contestaron la respuesta.

El primero de los préstamos anteriores hace referencia a créditos para comprar materias primas o materiales, pagar salarios; son créditos enfocados para financiar las necesidades del capital de trabajo, mayormente son otorgados al sector agropecuario. El segundo se refiere a préstamos dirigidos a financiar necesidades de inversión en activos fijos y fortalecer la infraestructura del negocio. Otras variables son la experiencia que han tenido los microempresarios en su negocio y en otros negocios. Por lo tanto se crearon dos variables, una de antigüedad donde esta presentada por los meses de experiencia que tuvo en otro negocio (cuando tiene 0 es que no tuvo experiencia) y la otra variable está representada por los meses que lleva con su negocio actual.

Por último, las cinco variables restantes son cuantitativas y se consideran importantes porque son las que están más relacionadas con el tema. Estas son el monto del préstamo que se les otorgó, la cantidad que pagarán mensualmente, las ganancias mensuales, las ventas mensuales todo esto medido en pesos, así como el tiempo que pagarán el crédito que esta medido en meses

La variable dependiente fue nombrada crédito, como se realizará una regresión LOGIT y además se está trabajando con el otorgamiento de préstamos, esta tenía que ser una variable dicotómica es decir que solo cuente con dos valores (0 y 1). Por lo tanto, las observaciones que tienen el número 1, significa que son personas que sí pueden pagar el préstamo que obtuvieron, los que tienen el número 0 son personas que no pueden pagar el préstamo. Con esto, al trabajar con neural tools podremos conocer las características que más influyen para que se pague un

crédito o no, además de pronosticar si una persona con determinadas características pueda cumplir con el pago. Para poder llegar a estos valores, lo que se hizo fue tomar las ventas mensuales que tenían los micronegocios, y restarles el pago mensual que tienen que hacer por el préstamo, el resultado que se obtuvo fue condicionado, si era mayor de \$8,344 o \$4,032, sí podía pagar el préstamo y por lo tanto era número 1 en la variable, de manera contraria, si era menor a las cantidades antes mencionadas no podía pagar el préstamo y se le asignaría el número 0.

Se trabaja con dos cantidades ya que Rodríguez y Aguilar (2013) consideran que para el 2012 las unidades de negocio que tienen ganancias mayores a los \$8,344 son capaces de solventar los gastos familiares, pero aquellos negocios que no generen ganancias mayores a \$4,032 no obtienen utilidades necesarias para cubrir el mínimo bienestar familiar. Por esta razón se trabajan con estas dos cantidades, teniendo las mismas variables que anteriormente se habían mencionado. En ambas variables se trabajan con un total de 4,489 datos históricos de microempresarios mexicanos.

**Tabla 1 Frecuencias**

	<b>Cumplen con el pago</b>	<b>No cumplen con el pago</b>	<b>Total</b>
<b>Variable \$4,032</b>	2691	1798	4,489
<b>Variable \$8,344</b>	1812	2677	4,489

En la siguiente Tabla 2 se pueden observar todas las variables independientes posibles a utilizar.

**Tabla 2 Variables independientes**

<b>VARIABLES DICOTÓMICAS</b>	
<ul style="list-style-type: none"> <li>NIVEDU: Nivel de educación</li> </ul>	1 Licenciatura o más, 0 No tiene licenciatura

<ul style="list-style-type: none"> <li>SEX: Sexo</li> </ul>	<p>1 hombres 0 mujeres</p>
<ul style="list-style-type: none"> <li>BC: Solicitó el préstamo en la Banca Comercial</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>BD: Solicitó el préstamo en la Banca de desarrollo</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>PG: Solicitó el préstamo en un programa de gobierno (1 verdadero, 0 falso)</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>PROV: Solicitó su préstamo en créditos a proveedores</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>CP: Solicitó su préstamo en cajas populares</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>PCI: Solicitó su crédito a través de prestamistas particulares</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>MANU: Negocio dedicado a la manufactura</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>COM: Negocio dedicado al comercio (1 verdadero, 0 falso)</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>SERV: Negocio dedicado a los servicios</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>PRES_R: Préstamos refaccionarios</li> </ul>	<p>1 verdadero 0 falso</p>
<ul style="list-style-type: none"> <li>PRES_H: Préstamos habilitación y avío</li> </ul>	<p>1 verdadero 0 falso</p>

<ul style="list-style-type: none"> <li>• EDO_CIV: Estado civil</li> </ul>	1 Con pareja 0 Sin pareja
<ul style="list-style-type: none"> <li>• (20-35): Porcentaje de pobreza del 20% al 35% en el estado</li> </ul>	1 verdadero 0 falso
<ul style="list-style-type: none"> <li>• (35-50): Porcentaje de pobreza del 35% al 50% en el estado</li> </ul>	1 verdadero 0 falso
<ul style="list-style-type: none"> <li>• (50-65): Porcentaje de pobreza del 50% al 65% en el estado</li> </ul>	1 verdadero 0 falso
<ul style="list-style-type: none"> <li>• (65-80): Porcentaje de pobreza del 65% al 80% en el estado (1 verdadero, 0 falso)</li> </ul>	1 verdadero 0 falso
<b>VARIABLES NUMÉRICAS</b>	
<ul style="list-style-type: none"> <li>• EDAD: Edad (años)</li> <li>• P: Cantidad del préstamo</li> <li>• MESESPRES: Número de meses en los que pagará el crédito</li> <li>• VENTAS: Ventas mensuales de los negocios. (pesos)</li> <li>• GANANCIAS: Ganancias mensuales de los negocios (pesos)</li> <li>• PAGOPRES: Cantidad que se paga mensualmente por el préstamo.</li> <li>• ACTUAL: Meses que lleva con el negocio</li> <li>• ANTIG: Meses de experiencia en otro negocio</li> </ul>	



## IV. Regresión Logit

Es importante mencionar que se está utilizando un modelo LOGIT debido a que la variable dependiente es dicotómica, además de que existen tres razones principales por las que no se puede utilizar un modelo de probabilidad lineal cuando se tiene la variable dependiente dicotómica. Primero, los términos de error son heteroscedásticos. Segundo, los errores no se distribuyen de manera normal porque la variable endógena solo puede tomar dos valores. Por último, las probabilidades que se predicen pueden ser mayores que 1 o menores que 0.

Dicho lo anterior, la regresión logística es un instrumento estadístico de análisis bivariado o multivariado, de uso tanto explicativo como predictivo. Resulta útil su empleo cuando se tiene una variable dependiente y un conjunto de  $m$  variables predictoras o independientes, que pueden ser cuantitativa o categóricas. (SEQC 2008). Este modelo es muy utilizado cuando lo que se quiere es conocer o pronosticar la probabilidad de que suceda o no algo. Además de ser una herramienta muy flexible en cuanto a la utilización de las variables independientes ya que pueden ser tanto cuantitativas como cualitativas (variables dummy).

### IV.1. Estimación del modelo Logit

Como se está trabajando con una gran cantidad de variables dicotómicas, lo que se decidió fue correr una por una las variables para saber si son significativas en relación con la variable dependiente, ya que si se corrían todas juntas era muy probable que la mayoría sean no significativas debido al problema de la multicolinealidad, que es cuando dos o más de las variables del modelo mantiene una relación lineal.

Hay que recordar que se están trabajando dos cantidades para condicionar la variable dependiente, que son \$4,032 y \$8,344, por lo tanto se tendrán dos diferentes variables dependientes, para cada una se correrán todas las variables que anteriormente se habían mencionado que son un total de 26. Con Eviews se

realiza la regresión de cada una de las variables con relación a la dependiente, seleccionando que sea logarítmica y se obtuvieron los siguientes resultados:

**Tabla 3 Resultados de las regresiones individuales \$4,032**

<b>SIGNIFICATIVAS INDIVIDUALMENTE CON \$4,032</b>	
<b>Dicotómicas</b>	<b>Numéricas</b>
<ul style="list-style-type: none"> <li>• Nivel de Educación</li> <li>• Giro: Manufactura</li> <li>• Giro: Comercio</li> <li>• Préstamo de habilitación y avío</li> <li>• Préstamo refaccionario</li> <li>• Estado Civil</li> <li>• Entidades Zona 1 Porcentaje de la población que sufre pobreza (65% - 80%)</li> <li>• Entidades Zona 2 Porcentaje de la población que sufre pobreza (35% - 50%)</li> <li>• Entidades Zona 3 Porcentaje de la población que sufre pobreza (20% - 35%)</li> </ul>	<ul style="list-style-type: none"> <li>• Cantidad del préstamo</li> <li>• Mensualidades del Préstamo</li> <li>• Ventas</li> <li>• Ganancias</li> <li>• Meses de que trabajo en otro negocio</li> <li>• Meses que lleva con el negocio actual</li> </ul>

**Tabla 4 Resultados de las regresiones individuales \$8,344**

<b>SIGNIFICATIVAS INDIVIDUALMENTE CON \$8,344</b>	
<b>Dicotómicas</b>	<b>Numéricas</b>
<ul style="list-style-type: none"> <li>• Nivel de Educación</li> <li>• Giro: Manufactura</li> <li>• Giro: Comercio</li> <li>• Préstamo de habilitación y avío</li> <li>• Préstamo refaccionario</li> <li>• Estado Civil</li> <li>• Entidades Zona 1 Porcentaje de la población que sufre pobreza (65% - 80%)</li> </ul>	<ul style="list-style-type: none"> <li>• Mensualidades del Préstamo</li> <li>• Ventas</li> <li>• Ganancias</li> <li>• Meses de que trabajo en otro negocio</li> <li>• Meses que lleva con el negocio actual</li> </ul>

<ul style="list-style-type: none"> <li>• Entidades Zona 2 Porcentaje de la población que sufre pobreza (35% - 50%)</li> <li>• Entidades Zona 3 Porcentaje de la población que sufre pobreza (20% - 35%)</li> </ul>	
--	--

De acuerdo a los resultados, y como podemos ver en las tablas anteriores (Tabla 3 y Tabla 4), las variables que fueron significativas con las dos cantidades son las mismas a diferencia que en la variable dependiente condicionada con \$4,032, la variable de cantidad del préstamo, fue significativa.

Posteriormente al conocer que estas variables individualmente fueron significativas, se realizó una sola regresión para significativas, se realizó una sola regresión para cantidad (\$4,032 y \$8,344) con sus variables significativas sus variables significativas respectivamente, teniendo en cuenta que era muy probable que el resultado fuera probable que el resultado fuera diferente y sean no significativas. El resultado se puede ver en las dos tablas puede ver en las dos tablas siguientes, la **Tabla 5** para la cantidad de \$4,032 y la

Tabla 6 para la cantidad d \$8,334.

**Tabla 5 Regresión \$4,032**

Dependent Variable: CRED4032  
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)  
Date: 06/20/16 Time: 16:26  
Sample: 1 4489  
Included observations: 4489  
Convergence achieved after 10 iterations  
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
NIVEDU	-0.240681	0.111262	-2.163190	0.0305
MANU	0.077348	0.137887	0.560952	0.5748
COM	-0.265565	0.101427	-2.618296	0.0088
PRES_H	0.359934	0.223534	1.610199	0.1074
PRES_R	0.195071	0.231045	0.844299	0.3985
EDO_CIV	-0.063440	0.091784	-0.691191	0.4894
P	3.42E-08	4.96E-08	0.689807	0.4903
MESES	0.000469	0.001327	0.353452	0.7237
VENTAS	0.000267	1.20E-05	22.36124	0.0000
GANAN	0.000144	1.94E-05	7.430422	0.0000
ANTIG	-0.000469	0.000516	-0.909276	0.3632
ACTUAL	-7.76E-05	0.000385	-0.201601	0.8402
_65_80_	0.013290	0.134939	0.098487	0.9215
_35_50_	-0.011129	0.107421	-0.103598	0.9175
_20_35_	-0.153683	0.125547	-1.224111	0.2209
C	-2.308164	0.257944	-8.948314	0.0000

McFadden R-squared	0.409764	Mean dependent var	0.599465
S.D. dependent var	0.490061	S.E. of regression	0.295244
Akaike info criterion	0.801855	Sum squared resid	389.9079
Schwarz criterion	0.824700	Log likelihood	-1783.763
Hannan-Quinn criter.	0.809905	Restr. deviance	6044.240
Restr. log likelihood	-3022.120	LR statistic	2476.714
Avg. log likelihood	-0.397363	Prob(LR statistic)	0.000000

Obs with Dep=0	1798	Total obs	4489
Obs with Dep=1	2691		

De acuerdo a los resultados, solo las variables ventas, ganancias y comercio fueron significativas teniendo menos de 0.05 de probabilidad. Las demás como se esperaba, ya que la mayoría son variables dicotómicas fueron no significativas, aunque es importante recordar que todas individualmente si lo fueron.

En una regresión lineal se cuenta con la  $R^2$ , para conocer qué tan bueno es el modelo, que se presenta, en esta caso para las regresiones no lineales se utiliza la  $R^2$  de Mc Fadden, la cual toma valores entre cero y así comprueba el grado de determinación de las variables explicativas, de tal manera que mientras más se acerca este del número 1, mejor será el modelo. En las regresiones que se presentan, el que se podría considerar mejor modelo es el que utiliza la variable \$8,344, aunque el resultado no es tan cercano a 1. Es muy frecuente que esta bondad de ajuste no tome valores muy altos, por lo que para los dos modelos se tiene un aceptable grado de determinación de las variables explicativas.

**Tabla 6 Regresión \$8,344**

Dependent Variable: CRED8344  
 Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)  
 Date: 06/20/16 Time: 17:23  
 Sample: 1 4489  
 Included observations: 4489  
 Convergence achieved after 9 iterations  
 Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
VENTAS	0.000290	1.08E-05	26.99220	0.0000
PRES_R	0.059299	0.266726	0.222324	0.8241
PRES_H	0.133829	0.261446	0.511879	0.6087
NIVEDU	-0.138688	0.126141	-1.099468	0.2716
MESES	0.001773	0.001524	1.163557	0.2446
MANU	-0.097682	0.157637	-0.619663	0.5355
GANAN	6.69E-05	1.58E-05	4.237473	0.0000
EDO_CIV	-0.129682	0.110927	-1.169074	0.2424
COM	-0.319263	0.119198	-2.678431	0.0074
ANTIG	-0.000947	0.000637	-1.485921	0.1373
ACTUAL	-0.000322	0.000465	-0.693650	0.4879
_65_80_	-0.000406	0.170405	-0.002381	0.9981
_20_35_	0.021497	0.147094	0.146145	0.8838
_35_50_	0.114275	0.128929	0.886338	0.3754
C	-3.698061	0.304036	-12.16322	0.0000
<hr/>				
McFadden R-squared	0.550411	Mean dependent var	0.403653	
S.D. dependent var	0.490684	S.E. of regression	0.240698	
Akaike info criterion	0.613147	Sum squared resid	259.2033	
Schwarz criterion	0.634564	Log likelihood	-1361.209	
Hannan-Quinn criter.	0.620695	Restr. deviance	6055.349	
Restr. log likelihood	-3027.674	LR statistic	3332.931	
Avg. log likelihood	-0.303232	Prob(LR statistic)	0.000000	
<hr/>				
Obs with Dep=0	2677	Total obs	4489	
Obs with Dep=1	1812			

En esta regresión al igual que la anterior las variables ventas, ganancias y comercio fueron significativas, mientras que las variables dicotómicas no lo fueron debido a que puede existir multicolinealidad.

Como se está trabajando con una regresión logit, hay que tener presente que los coeficientes que se obtiene no afectan de manera lineal, por lo que para interpretarlos se presenta el *odds ratio* de las variables independientes, al igual que la probabilidades que genera el modelo, el cual se presenta a continuación.

Recordando que el modelo logit representa una probabilidad  $P = \frac{1}{1+e^{-z}}$ , los coeficientes estimados en las regresiones nos servirán para conocer el valor de z.

Para la regresión \$4, 032

$$Z = 0.000267\text{Ventas} + 0.195071\text{Pres\_R} + 0,359934\text{Pres\_H} - 0,240681\text{Nivedu} + 0,000469\text{Meses} + 0,077348\text{Manu} + 0,000144\text{Ganan} - 0,06344\text{Edo\_Civ} - 0,265565\text{Com} - 0,000469\text{Antig} - 0,0000776\text{Actual} + 0,01329(65\_80) - 0,011129(35\_50) - 0,153683(20\_35) + 0,0000000342P$$

Para la regresión \$8, 344

$$Z = 0,000297\text{Ventas} + 0,059299\text{Pres\_R} + 0,133829\text{Pres\_H} - 0,138688\text{Nivedu} + 0,001773\text{Meses} - 0,097882\text{Manu} + 0,0000669\text{Ganan} - 0,12682\text{Edo\_Civ} - 0,319263\text{Com} - -0,000947\text{Antig} - 0,000322\text{Actual} - 0,000406(65\_80) + 0,114275(35\_50) + 0,021497(20\_35)$$

## IV.2. Interpretación de resultados

**Tabla 7 Impactos en la regresión \$4,032**

Variable	Coefficiente B	Odds Ratio	Impacto
VENTAS	0,000267	1,00026704	Positivo
PRES_R	0,195071	1,21539728	Positivo
PRES_H	0,359934	1,43323482	Positivo
NIVEDU	-0,240681	0,78609235	negativo
MESES	0,000469	1,00046911	Positivo
MANU	0,077348	1,080418	Positivo
GANAN	0,000144	1,00014401	Positivo
EDO_CIV	-0,06344	0,93853043	negativo
COM	-0,265565	0,7667726	negativo
ANTIG	-0,000469	0,99953111	negativo
ACTUAL	-0,0000776	0,9999224	negativo
65_80	0,01329	1,01337871	Positivo
35_50	-0,011129	0,9889327	negativo
20_35	-0,153683	0,85754382	negativo
P	3,42E-08	1,00000003	Positivo

### IV.2.1. Impactos

En la tabla anterior se presentan los impactos que tienen cada variable, ya sea de manera negativa o positiva, esto se determina de manera sencilla, si el coeficiente B es igual a cero nos dice que no afecta a la variable dependiente, si el coeficiente es negativo, nos dice que un aumento en esa variable afectara de manera

negativa en la variable dependiente, de manera contraria si el coeficiente es positivo, cuando tenga un aumento, también lo hará la variable dependiente.

En la regresión con \$4,032 podemos notar que todas las variables numéricas tuvieron un impacto positivo excepto el tiempo del negocio (ACTUAL) y la experiencia que haya tenido en otro negocio el microempresario (ANTIG). Por otro lado en referencia a las categóricas, ya que se manejan dos características en las mismas variables, se toma en cuenta la característica determinada con el número 1 para el impacto, por ejemplo la variable estado civil de la **¡Error! La autoreferencia al marcador no es válida.** tuvo un impacto negativo por lo tanto hace referencia a las microempresarios que tienen una pareja, de manera contraria los microempresarios que se encuentren solteros tendrán un impacto positivo en la variable dependiente.

**Tabla 8 impactos en la regresión 8,344**

Variable	Coeficiente	Odds Ratio	Impacto
VENTAS	0,00029	1,00029004	Positivo
PRES_R	0,059299	1,06109246	Positivo
PRES_H	0,133829	1,14319732	Positivo
NIVEDU	-0,138688	0,87049958	negativo
MESES	0,001773	1,00177457	Positivo
MANU	-0,097882	0,90675589	negativo
GANAN	0,0000669	1,0000669	Positivo
EDO_CIV	-0,12682	0,88089222	negativo
COM	-0,319263	0,72668441	negativo
ANTIG	-0,000947	0,99905345	negativo
ACTUAL	-0,000322	0,99967805	negativo
65_80	-0,000406	0,99959408	negativo
35_50	0,114275	1,12106037	positivo
20_35	0,021497	1,02172973	positivo

De la misma forma se detallan los impactos en la tabla 8, con la regresión \$8,344, teniendo como resultado de igual manera que todas las variables cuantitativas tuvieron un impacto positivo excepto las variables ANTIG y ACTUAL.

Si observamos las dos tablas anteriores (**¡Error! La autoreferencia al marcador no es válida.** y Tabla 8 ), aunque prácticamente son las mismas variables, podemos ver que no tienen el mismo impacto. Tal es el caso de cuatro variables, uno es el giro; manufactura, en la primera regresión tiene un impacto positivo y en la segunda el impacto es de manera contraria. Las tres variables restantes son las relacionadas a la entidad federativa, son tres de las cuatro zonas en las que se dividió el país, en la primera regresión con \$4,032 mostraba que un microempresario de la zona con más pobreza tenía un impacto positivo en la variable dependiente, es decir que los estados con menos pobreza iban a tener un impacto negativo. En la segunda regresión sucede lo contrario, los estados que se encuentren en las zonas con menos pobreza tendrán un impacto positivo, mientras que los estados con más pobreza afectaran de manera negativa.

#### IV.2.2. Odds ratio

De igual manera en la **¡Error! La autoreferencia al marcador no es válida.** y la Tabla 8, en la tercera columna se puede observar el *odds ratio*, que como se había mencionado antes es una manera de interpretar los coeficientes de la regresión logística. Este análisis de *odds ratio* se obtiene con el exponencial de los coeficientes y es el cociente de las probabilidades entre que ocurra un suceso respecto de que no ocurra. Su interpretación es la preferencia de una opción frente a otra, es decir el número de veces que es más probable que ocurra el fenómeno frente a que no ocurra. (Medina, 2003). Con el exponencial de los coeficientes obtenemos un razón *odds ratio*, tomando en cuenta que las demás variables permanecen iguales.

$$Odds Ratio = \frac{\rho(\acute{e}xito)}{\rho(fracaso)}$$

Podemos observar que en la variable ventas en la regresión \$4, 032 el *odds* hace referencia que al aumentar una unidad en las ventas la probabilidad de éxito es 1,00026704 veces mayor que la de fracaso, lo que significa que la probabilidad de pago incrementara cuando se tengan mayores ventas. Lo mismo sucede con la regresión \$8, 344 donde la probabilidad de éxito es 1,00029004 veces mayor.



Lo mismo sucede con las variables ganancias y meses, en la regresión \$4,032 el aumento de una unidad en las ganancias, hace que la probabilidad de éxito sea mayor 1.00014 veces, mientras que el plazo que se le otorga para pagar el crédito por cada que aumento de un mes, la probabilidad de éxito será mayor 1.00046 veces. Con la cantidad de \$8,344 la variable ganancias, dado el aumento de una unidad, la probabilidad de éxito será mayor 1.000066 veces y si aumenta el plazo del pago, por cada aumento de un es, la probabilidad de éxito es mayor 1.0017 veces. Y si solo para el caso de la regresión \$4,032, la variable préstamos, si aumenta una unidad, la probabilidad de éxito es mayor 1.00000003 veces, estas últimas variables mencionadas al igual que las ventas, como se había mencionado anteriormente tiene una relación positiva, dado un aumento en estas, la probabilidad de cumplir con el pago de un crédito será mayor.

En el caso de la antigüedad (experiencia que ha tenido en otro negocio) y el tiempo que llevan con el negocio que manejan (actual), en ambas regresiones el impacto fue negativo y al sacar el *odds ratio* el resultado muy similar. Por lo tanto se puede decir que la probabilidad de éxito es 0.99 veces, refiriendo a que es mayor la probabilidad de fracaso o incumplimiento del pago, por lo tanto los microempresarios que tengan experiencia y tengan mucho tiempo con su negocio tendrán menos probabilidad de pagar.

En el caso de las variables dicotómicas, como son relaciones entre dos resultados, e incluso entre varias variables, el análisis es un poco distinto, tal es el caso de los tipos de préstamos, en los resultados para la cantidad \$4,032, un microempresario que solicita un préstamo de habilitación y avío tiene más probabilidad de pagar que uno que solicita un crédito refaccionario, ya que el *odds ratio* es más alto. Lo mismo sucede con la cantidad de \$8,344, las personas que solicitan un préstamo de habilitación y avío tienen más probabilidades de pagar.

En cuanto al nivel de educación observamos que aproximadamente la relación es 3.1 a 4 = 0.78 para la regresión \$4, 032 lo que significa que por cada 3.1 microempresarios que tienen licenciatura y cumplen con el pago de crédito, hay 4 microempresarios que no cuentan con licenciatura y de igual manera son

cumplidores, para la regresión \$8,344 la relación es  $3.5 \text{ a } 4 = 0.87$  lo que quiere decir que una persona sin licenciatura es más probable que cumpla con el pago de un crédito. Lo mismo sucede con el estado civil, la relación es  $3 \text{ a } 3.2 = 0.93$  es decir por cada 3 microempresarios que tiene una pareja y cumplen con el pago de un crédito, 3.2 microempresarios sin pareja cumplen con el pago. Para la regresión \$8, 344 la relación es  $3.55 \text{ a } 4 = 0.88$ . Esto quiere decir que los microempresarios sin una pareja tendrán más probabilidades de pagar.

En relación al giro del negocio, se tienen dos variables que son comercio y manufactura, la que tiene más probabilidad de cumplir en la cantidad de \$4,032 es la manufactura ya que la relación es  $2.16 \text{ a } 2 = 1.08$  esto es que de 2.16 microempresarios que se dedican a la manufactura y cumplen con el pago de un crédito hay 2 microempresarios que no se dedican a este giro y cumplen con el pago; comparado con el caso del comercio que tiene una relación  $2 \text{ a } 2.6 = 0.76$  esto quiere decir que de 2 microempresarios que se dedican al comercio y cumplen con el pago de un crédito hay 2.6 que se dedican a otro giro y cumplen con el pago. Con esto comprobamos que un microempresario dedicado a la manufactura tendrá mayor probabilidad de pago que uno dedicado al comercio.

Para la regresión \$8,344, ambas variables tienen un *odds ratio* menor a uno lo que indica que los microempresarios que se dediquen al comercio o manufactura tendrán menos probabilidad de pago, pero el *odds ratio* del giro que es más cercano a uno o más grande es el de manufactura lo que indica que su probabilidad de incumplimiento será menor. Las razones están dados para manufactura es de  $2.72 \text{ a } 3 = 0.90$  y del comercio es de  $2.18 \text{ a } 3 = 0.726$

Por último, las variables que hacen referencia a los estados, que se encuentren en diferentes clasificaciones, con la cantidad \$4,032, los estados que tienen entre 65 y 80% de pobreza tienen una razón  $2.02 \text{ a } 2 = 1.013$  aunque es muy pequeña la diferencia, esta clasificación significa que los microempresarios que sean de la clasificación de los estados donde hay más pobreza son los que tienen mayor probabilidad de pagar, ya que las otras dos clasificaciones restantes tuvieron *odds ratio* menores a uno, por lo tanto si los microempresarios que se encuentren en los

estados de estas dos clasificaciones tendrán menos probabilidad de cumplir con el pago de un crédito. Con la regresión \$8,344 la clasificación con 35 a 50 % de pobreza en los estados es la que tiene mayor probabilidad de cumplimiento ya que su razón es 2.43 microempresarios que están en un estado de esta clasificaciones que cumplen a 2 microempresarios que no se encuentran en esta clasificación y cumplen con el pago. Ser de los estados que están en la clasificación de 20 a 35% también aumentara la probabilidad de cumplimiento aunque en menor medida y ser de los estados que se encuentran en la clasificación 65 a 80% hará que disminuya esta probabilidad.

#### IV.2.3. Probabilidades con el modelo Logit

Retomando el modelo que se obtuvo y se señaló al final del apartado de estimación del modelo:

$$P = \frac{1}{1 + e^{-z}}$$

Donde **z** es:

Para la regresión \$4,032

$$z = 0.000267\text{Ventas} + 0.195071\text{Pres\_R} + 0.359934\text{Pres\_H} - 0.240681\text{Nivedu} + 0.000469\text{Meses} + 0.077348\text{Manu} + 0.000144\text{Ganan} - 0.06344\text{Edo\_Civ} - 0.265565\text{Com} - 0.000469\text{Antig} - 0.0000776\text{Actual} + 0.01329(65\_80) - 0.011129(35\_50) - 0.153683(20\_35) + 0.000000342P$$

Para la regresión \$8,344

$$z = 0.000297\text{Ventas} + 0.059299\text{Pres\_R} + 0.133829\text{Pres\_H} - 0.138688\text{Nivedu} + 0.001773\text{Meses} - 0.097882\text{Manu} + 0.0000669\text{Ganan} - 0.12682\text{Edo\_Civ} - 0.319263\text{Com} - -0.000947\text{Antig} - 0.000322\text{Actual} - 0.000406(65\_80) + 0.114275(35\_50) + 0.021497(20\_35)$$

Con lo anterior podemos ver a **z** de manera lineal y los coeficientes nos muestran un cambio en la unidad de las variables pero de manera logística, esto toma sentido cuando **z** es puesta en la función logística y obtenemos una probabilidad (valores de 0 a 1) dadas las variables anteriores.

Por lo tanto para probar este modelo y las probabilidades que otorga, trabajaremos con una persona X que tendrá las siguientes características:

Para la regresión \$4,344

- Microempresario (hombre)
- Casado
- Del estado de Quintana Roo (clasificación 35 a 50%)
- Con licenciatura
- Dedicado al comercio
- Solicitando crédito refaccionario
- Solicitando un crédito de \$50,000,
- Plazo para pagar el crédito de 24 meses
- Ventas mensuales de \$12 000
- Ganancias mensuales de \$8,400
- Dueño por 60 meses de su negocio (5 años)
- 10 meses de experiencia en otro negocio

**Tabla 9** Probabilidad \$4,032

$\alpha$	-2,308164		
VARIABLES	$\beta$	$x$	$\beta x$
VENTAS	0,000267	12000	3,204
PRES_R	0,195071	1	0,195071
PRES_H	0,359934	0	0
NIVEDU	-0,240681	1	-0,240681
MESES	0,000469	24	0,011256
MANU	0,077348	0	0
GANAN	0,000144	8400	1,2096
EDO_CIV	-0,06344	1	-0,06344
COM	-0,265565	1	-0,265565
ANTIG	-0,000469	10	-0,00469
ACTUAL	-0,0000776	60	-0,004656
65_80	0,01329	0	0
35_50	-0,011129	1	-0,011129
20_35	-0,153683	0	0
P	3,42E-08	50000	0,00171
$\alpha + \beta_1x_1 + \dots + \beta_nx_n = Z$			1,723312

$$P = \frac{1}{1 + e^{-1,723312}} = .8485$$

Con estas características la probabilidad de cumplimiento de pago es de **84.85%**

Para la regresión \$8,032

- Microempresario (hombre)
- Casado
- Del estado de Quintana Roo
- Con licenciatura
- Dedicado al comercio
- Solicitando crédito refaccionario
- Plazo para pagar el crédito de 24 meses
- Ventas mensuales de \$12 000
- Ganancias mensuales de \$8,400
- Dueño por 60 meses de su negocio (5 años)
- 10 meses de experiencia en otro negocio

**Tabla 10** Probabilidades \$8,344

$\alpha$	-3,698061		
Variables	$\beta$	$\chi$	$\beta\chi$
<b>VENTAS</b>	0,00029	12000	3,48
<b>PRES_R</b>	0,059299	1	0,059299
<b>PRES_H</b>	0,133829	0	0
<b>NIVEDU</b>	-0,138688	1	-0,138688
<b>MESES</b>	0,001773	24	0,042552
<b>MANU</b>	-0,097882	0	0
<b>GANAN</b>	0,0000669	8400	0,56196
<b>EDO_CIV</b>	-0,12682	1	-0,12682
<b>COM</b>	-0,319263	1	-0,319263
<b>ANTIG</b>	-0,000947	10	-0,00947
<b>ACTUAL</b>	-0,000322	60	-0,01932
<b>65_80</b>	-0,000406	0	0
<b>35_50</b>	0,114275	1	0,114275
<b>20_35</b>	0,021497	0	0
<b><math>\alpha + \beta_1\chi_1 + \beta_i\chi_i = Z</math></b>			-0,053536

$$P = \frac{1}{1 + e^{-(-0,053536)}} = .4866$$

Con estas características la probabilidad de cumplimiento de pago es de **48.66%**

Para obtener las probabilidades anteriores lo único que se realizó fue sustituir en **z** los nombres de las variables por los valores correspondientes y después colocar este en la función logística como se presentó en la Tabla 9 y Tabla 10.

## ***V. Neuraltools***

### **V.1. Aplicación de *Neuraltools***

Para trabajar los datos con el programa *Neuraltools*, existe una serie de pasos a realizar. El primero es revisar y ordenar los datos que se van a utilizar, teniendo muy claro lo que se pretende analizar, en esta caso se está trabajando con préstamos por lo que queremos conocer si se le otorga un préstamo a una persona con ciertas características en base a datos históricos.

El siguiente paso es crear con los datos ordenados un nuevo grupo de datos en *Neuraltools*, teniendo claro cuáles son las variables independientes y la variable dependiente, para asignarle una clasificación ya sea numérica o cualitativa o en dado caso la opción “no usar” cuando no se necesite utilizar la variable.

Después de tener todas las variables con su debida configuración, se selecciona entrenar en el menú de herramientas de *Neuraltools* activando las opciones de probar casos aleatoriamente con el porcentaje que se desee, el programa trae con un 20% que es el recomendado a utilizar. En el caso que hagan falta valores en la variable dependiente, se selecciona la opción predecir los valores faltantes. De igual manera es muy recomendable activar la opción de predicción en vivo ya que al realizar algún cambio en las variables independientes automáticamente cambian los resultados. Y por último existe una opción de selección de un algoritmo, donde se selecciona una red y hay tres opciones: búsqueda de mejor red, red PN/GRN y Red MLF. La más utilizada es la segunda para problemas de clasificación ya que es un algoritmo de red neuronal probabilístico rápido.

Al tener toda esta configuración se selecciona la pestaña siguiente y luego la pestaña entrenar. Obteniendo como resultado un resumen, de lo que realizo

*Neuraltools*. En este se detallan de los entrenamientos, pruebas y predicciones que se llevaron a cabo con los datos. Mostrando como resultado lo que se realizó ya se una predicción, una prueba o un entrenamiento; en las pruebas y en las predicciones se muestra la predicción del programa de acuerdo a los valores que puede tomar la variable dependiente además del porcentaje de predicción que sea correcta y el porcentaje de que esto sea incorrecto. De acuerdo al porcentaje más alto se selecciona si el resultado es correcto o incorrecto.

## V.2. Aplicación *Neuraltools* a los datos de los microempresarios.

Como se había mencionado antes, lo primero que se tiene que hacer al utilizar este programa es acomodar las variables, para este trabajo se utilizaran dos hojas de Excel. En una la variable dependiente está condicionada a \$4,032 y la otra a \$8,344. En ambas se trabajan las mismas variables.

Como se tenían veintiséis variables independientes, se realizó la regresión logística para saber cuáles eran más significativas, y así trabajar con ellas en *Neuraltools*. Por lo tanto todas las variables significativas que se presentaron en las regresiones fueron utilizadas, pero además se agregaron variables que no fueron significativas pero que era importante trabajar con ellas. Tal es el caso de la edad, del sexo, y de una de las zonas en las que está dividido el país. En el caso de la regresión \$8,344 no fue significativa la variable préstamo pero se consideró importante tomar en cuenta. En la siguiente Tabla 11 se presentan las variables que se utilizaron en ambas hojas de Excel, mostrando su respectiva clasificación de acuerdo a *Neuraltools*.

**Tabla 11** Variables *Neuraltools*

<b>Variable</b>	<b>Clasificación</b>
<b>(20-35)</b>	Categórica independiente
<b>(32-50)</b>	Categórica independiente
<b>(50-65)</b>	Categórica independiente
<b>(65-80)</b>	Categórica independiente
<b>NIVEDU</b>	Categórica independiente

<b>EDAD</b>	Numérica independiente
<b>MANU</b>	Categórica independiente
<b>COM</b>	Categórica independiente
<b>SERV</b>	Categórica independiente
<b>PRES_H</b>	Categórica independiente
<b>PRES_R</b>	Categórica independiente
<b>EDO_CIV</b>	Categórica independiente
<b>SEX</b>	Categórica independiente
<b>P</b>	Numérica independiente
<b>MENSP</b>	Numérica independiente
<b>MESES</b>	Numérica independiente
<b>VENTAS</b>	Numérica independiente
<b>GANAN</b>	Numérica independiente
<b>ANTIG</b>	Numérica independiente
<b>ACTUAL</b>	Numérica independiente
<b>CRED8344</b>	Categórica dependiente

Ya que esta creado el conjunto de datos en el programa, se procede a entrenar las variables, seleccionando la red PN/GRN que es la más apta para este tipo de situaciones, ya que se trata de una red neuronal probabilística. Como igual lo que se quiere es predecir y probar los datos que se entrena, se selecciona la opción probar el 20% del total de los datos. Este procedimiento que realiza *Neuraltools* puede llevar unos cuantos minutos. Por último el programa arroja el resultado de la predicción o las predicciones que se querían conocer, además de dar un resumen detallado del entrenamiento, prueba y predicción. De igual manera en el resumen se detalla el impacto que las variables tuvieron, que es lo que más nos interesa para conocer las características que más afectan en el pago de un préstamo.

Como el programa *Neuraltools* realiza predicciones, para una situación en particular de acuerdo a los datos históricos, para este caso se decidió trabajar con



el mismo microempresario X que se utilizó para obtener las probabilidades con el modelo logit pero aumentando ciertas características con las que no se trabajaron en ese apartado.

Para ambas cantidades (\$4,032 y \$8,344) se utilizaran las mismas características:

- Microempresario (hombre)
- 40 años
- Casado
- Del estado de Quintana Roo (clasificación 35 a 50%)
- Con licenciatura
- Dedicado al comercio
- Solicitando crédito refaccionario
- Solicitando un crédito de \$50,000,
- Plazo para pagar el crédito de 24 meses
- Mensualidades de \$2,708
- Ventas mensuales de \$12 000
- Ganancias mensuales de \$8,400
- Dueño por 60 meses de su negocio (5 años)
- 10 meses de experiencia en otro negocio

Para la cantidad de \$4,032 la predicción que se obtuvo fue 1, es decir que la persona con las características que se consideraron anteriormente cumplirá con el pago del crédito, teniendo un porcentaje de acierto de 96.03%.

Con los datos de la cantidad de \$8,344 el resultado que se obtuvo fue de igual manera 1, lo que significa que el microempresario que cuente con las características que se mencionaron anteriormente cumplirá con el pago, teniendo un porcentaje de acierto de 85.17%.

Como lo que se quiere conocer es que características en los microempresarios mexicanos influyen en que cumplan o no con el pago de un crédito, se decidió cambiar una a una las variables para conocer el impacto que podrían tener estas en la variable dependiente y en cuanto se modifica el resultado inicial. Por lo tanto en la Tabla 12 y Tabla 13 se presentan el resultado de lo que sucedió. Posteriormente, se presentan las gráficas del impacto que *neuraltools* considero para cada variable.

**Tabla 12** Cambio en las variables en \$4,032

<b>NEURALTOOLS 4032</b>		
<b>Variables</b>	<b>Variable dependiente</b>	<b>Porcentaje de acierto</b>
Características iniciales	1	95.56%
Mujer	1	94.61%
Sin licenciatura	1	No tiene efecto
Sin pareja	1	95.29%
Zona 1 (20-35)	1	94.48%
Zona 3 (50-65)	1	95.71%
Zona 4 (65-80)	1	95.78%
Servicios	1	96.03%
Manufactura	1	95.88%
Préstamo de habilitación o avío	1	96.75%
Ganancias \$6 800	1	88.89%
Más joven (30 años)	1	96.02%
Cantidad de préstamo igual, pero Disminuye el plazo a 10 meses y mensualidades de 4584	1	96.04%
Ventas de \$8 000	1	78.31%
Ventas de \$5 000	0	69.51%
Ventas de \$20 000	1	99.28%
Antigüedad 24 meses	1	No tiene impacto

Dueño por 96 meses (8 años)	1	94.04%
-----------------------------	---	--------

En la tabla anterior (Tabla 12) que hace referencia a las variables de la cantidad \$4,032, se puede observar que los cambios al modificar una variable son muy pequeños, pero en ellos podemos apreciar de qué manera afecta a la variable dependiente, ya sea de manera positiva o negativa.

Las variables que al modificar, tuvieron un aumento en la probabilidad de acierto, fueron cuatro; la primera es la que hace referencia al tipo de préstamo, cuando se trata de un préstamo de habilitación o avío, es más probable que se cumpla con el pago del crédito que cuando se pide un préstamo refaccionario ya que el primero tiene 0.72% más de probabilidad. El segundo cambio donde se tuvo un aumento fue cuando el plazo del crédito se hace más corto (10 meses) y las mensualidades de este aumentan, teniendo 0.01% más de probabilidad de acierto.

El tercer cambio fue en la variable ventas, ya que como se esperaba si estas aumentan, es más probable que se cumpla con el pago, por lo tanto si el microempresario con las características principales ahora tiene ventas de \$20 000, la probabilidad de acierto ascenderá a 99.28%, es decir que es casi imposible que no cumpla con el pago. Por lo tanto si las ventas disminuyen, la probabilidad de acierto también lo hará, en la Tabla 12 podemos observar que si las ventas llegan a \$5,000 pesos, la predicción cambia a 0 (no cumple con el pago) con una probabilidad de acierto de 69.51%. Por último el cuarto cambio, fue en el aumento de los años que lleva el microempresario con su negocio, si en lugar de cinco años fueran ocho, se tendría un 0.01% más de probabilidad de acierto.

Los cambios restantes, tuvieron un impacto negativo en la probabilidad de acierto, como es el caso de las clasificaciones de los estados, donde de manera inesperada, los estados donde se cuenta con mayor pobreza fue donde menos disminuyo la probabilidad de acierto, por lo tanto los estados que tienen más probabilidad de cumplir con el pago son lo que tienen un porcentaje de pobreza

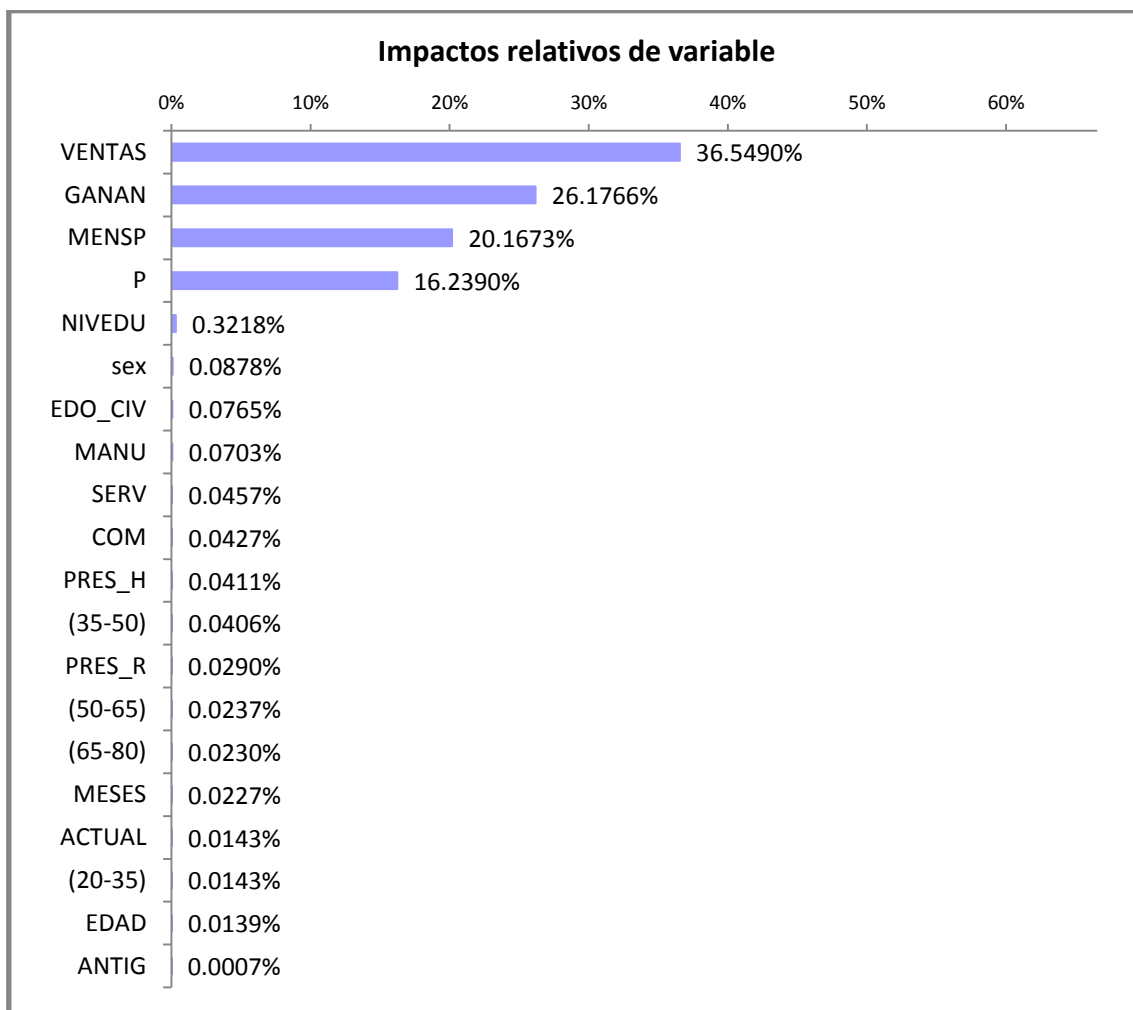
entre 35 y 50% y los que tienen menos probabilidad son los que tienen un porcentaje entre 20 y 35%.

Otros resultados fueron que las mujeres tienen menos probabilidad de acierto (más de 2% menos), al igual que las personas que no tengan una pareja (menos de 1%) y sean más jóvenes (diferencia de 0.01%). En relación al giro del negocio, se puede observar que cuando se presta algo tipo de servicio es cuando mayor probabilidad de acierto existe, seguida de la manufactura y luego el comercio, aunque la diferencia entre estos es mínima.

Las ganancias al ser disminuidas, como se esperaba, la probabilidad de acierto se hizo más pequeña, quedando en 88.89% disminuyendo un 7.14 en la probabilidad de acierto. Dos cambios, no tuvieron efecto en la variable dependiente los cuales son el nivel de educación y la experiencia que hayan tenido en otro negocio.

En la gráfica siguiente se presentan, los impactos de cada variable según *Neuraltools*. Esto es en relación, al impacto que tuvo cada variable sobre la variable dependiente en la predicción. Se puede observar que las variables que más impacto tienen son las cuantitativas, la primera de ellas como se esperaba son las ventas con un 36.54% seguido de las ganancias con un 26.17% y de la mensualidad que se pagó por el crédito con un 20.16%. Las variables cualitativas al ser varias, cada una tiene un mínimo impacto sobre la variable dependiente. La que mayor impacto tiene es el nivel de educación con 0.321% y la que menor impacto tiene es la experiencia que se tenga en otro negocio con un 0.0007%.

**Gráfica 1 Impacto de las variables \$4,032**



Lo anterior fue en relación a los datos de la cantidad de \$4,032, por lo que a continuación se presentan los resultados obtenidos con la cantidad de \$8,344.

**Tabla 13 Cambio en las variables de \$8,344**

<b>NEURALTOOLS 8344</b>		
<b>Variables</b>	<b>Variable dependiente</b>	<b>Porcentaje de acierto</b>
Características iniciales	1	84.88%
Mujer	1	85.30%
Sin licenciatura	1	85.15%

Sin pareja	1	85.21%
Zona 1 (20-35)	1	85.03%
Zona 3 (50-65)	1	85.38%
Zona 4 (65-80)	1	85.51%
Servicios	1	85.17%
Manufactura	1	85.53%
Préstamo de habilitación o avío	1	86.32%
Ganancias \$6 800	1	83.45%
Más joven 30 años	1	85.23%
Cantidad de préstamo igual, pero Disminuye el plazo a 10 meses y mensualidades de 4584	1	85.19%
Ventas de \$8 000	0	97.77%
Ventas de \$5 000	0	94.88%
Ventas de \$20 000	1	100%
Antigüedad 24 meses	1	85.15%
Dueño por 96 meses (8 años)	1	85.14%

Al igual que con la cantidad de \$4,032 los cambios que se dieron fueron muy pequeños, pero en esta ocasión la mayoría de los cambios en las variables hizo que aumentara la probabilidad de acierto. Recordando, las características iniciales dieron como respuesta 1 (si cumple con el pago), con una probabilidad de acierto de 85.17%.

Con estos datos, al realizar el cambio en el sexo, la probabilidad de acierto ascendió a 85.30% es decir que las mujeres tienen más probabilidad de pagar. Lo mismo sucedió con el sexo, ya que al no tener pareja se aumenta la probabilidad

de acierto. En cuanto a la clasificación de los estados, donde se tiene mayor probabilidad de acierto es con los estados que tiene un porcentaje de pobreza entre 65% y 80%, lo que quiere decir que los “pobres” son los que mayor probabilidad de pago tienen.

En relación al giro del negocio, el que mayor probabilidad de acierto tuvo fue el de la manufactura con un 85.53% y el de menos acierto fue los comerciantes con 84.88%. Sobre la edad al ser más jóvenes se tendrá un aumento de la probabilidad. En relación al tipo de préstamo habilitación o avío tuvo 1.25% mayor probabilidad de acierto de la variable dependiente, la cual seguía siendo 1.

Se puede observar que las ventas tienen un impacto positivo, ya que al aumentar la cantidad, la probabilidad de pago iba a incrementar. En el apartado de las ventas, se puede observar que si se cambia a \$8,000 la variable dependiente iba a cambiar a 0 (no cumple con el pago) teniendo un probabilidad de 97.77% y si las ventas bajaran a un más a \$5 000, la variable dependiente seguiría siendo 0 pero ahora con una probabilidad de acierto del 100%. Sucede lo mismo con la variable ganancias al disminuir estas a \$6,800 como se esperaba la probabilidad también se hizo menor quedando con 83.45%

Si el préstamo disminuye su plazo y por ende su mensualidad incrementa, esto hará que se tenga un aumento en la probabilidad de incumplimiento. Por último, los cambios en el aumento de la experiencia que se tenga en otro negocio, y el tiempo con el negocio actual provocaron un impacto negativo en la variable dependiente.

El impacto que cada una de las variables tuvo sobre la variable dependiente, de acuerdo a *Neuraltools*, se presentan a continuación. Se puede observar que tal como paso con la cantidad de \$4,032, las variables cuantitativas fueron las que más impacto tuvieron. Aunque con estos datos analizados, la variable que más afecto fue la de ganancias con un 31.70%, seguida de esta, es la cantidad que se pide por un préstamo con un 27.64% y con una muy pequeña diferencia siguen las ventas con un 26.47%. La variable cuantitativa con menor efecto fue la edad de los microempresarios con un 0.12% de impacto.

Las variables cualitativas al ser varias, tuvieron el mínimo impacto sobre la variable dependiente, la que mayor afectación tuvo fue el sexo del microempresario con un 0.58% y la que menor impacto tuvo fue el grupo de estados donde el porcentaje de pobreza es de 35% a 50% teniendo 0.0022 % de impacto.

**Gráfica 2** Impacto de las variables \$8344





## VI. Discusión de resultados

El uso de las herramientas de *Neuraltools* y la regresión logística es muy útil para cuando se quiere trabajar con modelos de calificación crediticia, además de que se son muy sencillos de utilizar. En esta trabajo lo que se quería conocer son las características que afectan a los microempresarios mexicanos para que cumplan con el pago de un crédito o no. Esto sería de mucha utilidad para los bancos o instituciones que se dedican a otorgar préstamos, ya que muchas veces existe la incertidumbre sobre si una persona al otorgarle un préstamo, va a cumplir con sus obligaciones. Provocando que muchas veces se nieguen créditos a personas que si cumplirían y de manera contraria otorgando créditos a personas que no cumplirían con el pago, generando pérdidas para la institución o la empresa.

En ambas herramientas se obtuvieron resultados parecidos, por lo que a continuación se hace un análisis de los resultados importantes que se tuvieron con la regresión logística y *Neuraltools*.

### VI.1. Regresión logística

La regresión logística es muy útil cuando se quiere conocer la probabilidad de algún suceso, y cuando se utilizan variables dicotómicas ya sea independiente o dependiente, por eso para este trabajo se decidió realizar esta regresión. De acuerdo a los resultados algunas de las variables tuvieron el impacto que se esperaba, tal es el caso de la ventas y las ganancias que cuanto mayor sean estas, mayor probabilidad se tendrá de cumplir con el pago del crédito. Otra variable es el plazo que se da para pagar el crédito, donde a mayor plazo, las mensualidades pueden reducirse un poco y es más probable que los microempresarios cumplan con el pago.

Otras como las clasificaciones de los estados, llaman mucho la atención, ya que hacen referencia a que en los estados donde mayor pobreza existe es donde los microempresarios mexicanos tienen mayor probabilidad de pagar (esto cuando se utiliza la cantidad de \$4,032); lo lógico es pensar que los estados donde existe menos pobreza es donde mayor probabilidad de pago debieran tener, lo que lo

hace un tema interesante por analizar. Aunque de manera contraria cuando se utiliza la cantidad de \$8,344 esta relación se invierte, siendo los estados con menos pobreza los que tienen mayor probabilidad de pagar.

Otras variables que dieron un resultado distinto al esperado, fueron la experiencia que se haya tenido en otro negocio y el tiempo que se tiene con el negocio actual. Se esperaba que mientras más experiencia, mejores utilidades deben de haber y por lo tanto debe existir una relación positiva con el pago del crédito, pero en los dos grupos de datos analizados, esta relación fue negativa, es decir que los nuevos emprendedores son los que mayor probabilidad de pagar tienen.

En relación a los tipos de préstamos, Rodríguez (2015) demuestra que las personas que solicitan préstamos refaccionarios son las que mayor probabilidad tienen de pagar, pero en los resultados obtenidos para este trabajo se demuestra lo contrario. Por lo tanto a manera de comparación se realizaron dos regresiones para cada cantidad con ambas variables y el resultado fue el contrario y como se esperaba, la variable más significativa fue la de los créditos refaccionarios.

El nivel de educación igual llama mucho la atención, ya que los que mayor probabilidad tienen de pagar son los microempresarios que no cuentan con licenciatura, cuando lo que se piensa es que alguien que esté más preparado con estudios genera mayores utilidades para su negocio y cumpla con el pago del crédito. Y por último otro resultado importante es que los microempresarios que no tienen pareja son los que mayor probabilidad tienen de pagar, se podría pensar que es debido a que tienen menos gastos, al no vivir con otra persona.

En cuanto a la aplicación del modelo es muy interesante observar y analizar los resultados ya que las probabilidades que se obtuvieron fueron muy diferentes, cuando los coeficientes eran muy parecidos y lo único distinto en estas era la variable Préstamos.

## VI.2. Neuraltools

Neuraltools es una herramienta mucho más sencilla de aplicar en comparación de la regresión logística. Es muy útil para los temas relacionados al otorgamiento de los créditos, ya que cuenta con un sistema predictorio en base a datos históricos. Además de dar un resultado de los impacto de las variables que se utilizan y el porcentaje de acierto que se tiene en relación a la predicción estimada. Al igual que la regresión logit, *Neuraltools* permite utilizar variables dicotómicas y no solo numéricas.

Los resultados fueron parecidos a los obtenidos con las regresiones en relación a las variables que se utilizaron en ambos procedimientos. Tal es el caso de las ventas que se demuestra que al aumentarlas la probabilidad del cumplimiento del pago también aumenta. Otro caso es el de los créditos habilitación o avío donde en ambas grupos de datos, se demostraba (\$4,032 y \$8,344) que los que solicitaran este tipo de préstamo tendría una mayor probabilidad de pago.

Un resultado que llamo la atención en el conjunto de datos de \$8344, fue que al disminuir las ganancias, la probabilidad aumento, cuando se esperaba que tuviera una relación positiva.

En cuanto a la clasificación de los estados, en ambos grupos de datos, se vuelve a demostrar que los estados con menor pobreza son los que mayor probabilidad de incumplimiento tienen. De igual manera en ambos grupos, se demuestra que las mujeres son las que mayor probabilidad de pago tienen, aunque la diferencia con los hombres es mínima, tal como lo demuestran Rodríguez y Hernández (2013) en una investigación llamada matriz de probabilidad de transición de microcréditos; el caso de una microfinanciera mexicana donde mencionan que el cumplimiento del pago no es cuestión de género, si no de responsabilidad.

En cuanto a la edad, que podría considerarse otro factor importante, en el primer conjunto de datos (\$4,032) se demuestra que a mayor edad menor probabilidad de pago y en el segundo conjunto (\$8,344) el resultado es de manera contraria, en ambos resultados el cambio en la probabilidad de incumplimiento es mínimo. Otra

variable que tuvo dos resultados distintos fue la de estado civil, donde los datos trabajados con la cantidad de \$4,032, fueron iguales a los que se obtuvieron en la regresión, ya que demuestra que microempresarios solteros tendrían mayores probabilidades de pago, resultando de manera contraria en el conjunto de datos de \$8,344. Esto podría deberse a que en este último se tienen mayores ingresos por lo que no afecta en el pago si se tiene una pareja o no, ya que se tiene más dinero para gastar.

Las variables restantes no presentan cambios significativos o que se consideren importantes. Para terminar hay que tener presente, que al trabajar con esta herramienta, los datos no siempre serán los mismo, pero sí muy parecidos. Ya que los entrenamientos y pruebas que realiza son de manera aleatoria. Por lo tanto los datos analizados son en base a una red neuronal específica, que probablemente no pueda volver a suceder.

### VI.3. Comparación de resultados

Como se había mencionado antes, los resultados obtenidos en ambos modelos de *credit scoring* fueron muy similares. Por lo tanto se realizó una tabla de comparación entre los resultados de la regresión logística y las redes neuronales.

**Tabla 14** Comparación de impactos \$4,032

<b>Variable</b>	<b>Regresión Logit</b>	<b>Redes Neuronales</b>
<b>(20-35)</b>	negativo	negativo
<b>(35-50)</b>	<u>negativo</u>	<u>positivo</u>
<b>(50-65)</b>	-	positivo
<b>(65-80)</b>	positivo	positivo
<b>NIVEDU</b>	negativo	sin efecto
<b>EDAD</b>	-	positivo
<b>MANU</b>	positivo	positivo
<b>COM</b>	negativo	negativo
<b>SERV</b>		positivo

<b>PRES_H</b>	positivo	positivo
<b>PRES_R</b>	positivo	positivo
<b>EDO_CIV</b>	<u>negativo</u>	<u>positivo</u>
<b>P</b>	positivo	positivo
<b>MENSP</b>	-	positivo
<b>MESES</b>	positivo	negativo
<b>VENTAS</b>	positivo	positivo
<b>GANAN</b>	positivo	positivo
<b>ANTIG</b>	negativo	sin efecto
<b>ACTUAL</b>	negativo	negativo
<b>SEX</b>	-	positivo

En la Tabla 14 podemos observar las comparaciones entre los resultados que se obtuvieron en la regresión logística y las redes neuronales cuando se trabajo con la cantidad de \$4,032. Como se puede observar con las redes neuronales se trabajaron más variables, por lo tanto la comparación es en base a las utilizadas en la regresión logística. En ambos métodos los resultados fueron muy similares, solo en tres variables se tuvo un resultado contrario en la zona geográfica (35-50) donde la regresión logística tuvo un impacto negativo y en las redes neuronales un impacto positivo, aunque en esta última (redes neuronales) la variable fue una característica inicial, y al compararla con las otras zonas, su probabilidad no fue la más alta aunque tampoco la más baja, entonces igual se podría considerar como un impacto negativo en relación a dos zonas geográficas. La segunda variable fue el estado civil, en la regresión logística se tuvo un impacto negativo, es decir que los microempresarios sin pareja tenían más probabilidad de pagar, por el contrario en las redes neuronales los microempresarios con pareja es más probable que paguen. Al realizar un análisis más detallado (en base a la características iniciales que se mencionaron anteriormente en los apartados IV y V) se puede observar que el impacto negativo (logit) en relación a las probabilidades es de 0.80% (las microempresarios sin pareja tendrán 0.80% más probabilidad de pagar). Con

respecto al impacto positivo (redes neuronales), las microempresarios con pareja tendrán 0.27% más probabilidad de pagar, comparando ambas el cambio más significativo se da en el impacto negativo. Y por último la variable relacionada con los plazos en los que se otorga el crédito (meses), en la regresión logit se tuvo un impacto positivo es decir mientras mayor sea el plazo más probabilidad de pagar, por el contrario con las redes neuronales si el plazo es menor se tiene una mayor probabilidad de pagar, en ambos el cambio en la probabilidad fue mínimo aunque el más significativo se dio en las redes neuronales.

En la Tabla 15 se presenta la comparación de los dos modelos con la cantidad \$8,344. Como se puede observar para esta cantidad cinco variables obtuvieron resultados contrarios.

**Tabla 15 Comparación de impactos \$8,344**

<b>Variable</b>	<b>Logit</b>	<b>Redes neuronales</b>
<b>(20-35)</b>	positivo	positivo
<b>(35-50)</b>	positivo	<u>negativo</u>
<b>(50-65)</b>	-	positivo
<b>(65-80)</b>	negativo	<u>positivo</u>
<b>NIVEDU</b>	negativo	negativo
<b>EDAD</b>	-	negativo
<b>MANU</b>	negativo	<u>positivo</u>
<b>COM</b>	negativo	negativo
<b>SERV</b>	-	positivo
<b>PRES_H</b>	positivo	positivo
<b>PRES_R</b>	positivo	positivo
<b>EDO_CIV</b>	negativo	negativo
<b>P</b>	-	positivo
<b>MENSP</b>	-	positivo
<b>MESES</b>	positivo	positivo
<b>VENTAS</b>	positivo	positivo

<b>GANAN</b>	positivo	positivo
<b>ANTIG</b>	negativo	<u>positivo</u>
<b>ACTUAL</b>	negativo	<u>positivo</u>
<b>SEX</b>	-	negativo

La primera variable es la zona geográfica (35-50), que obtuvo como resultado un impacto positivo en la regresión logística y en las redes neuronales un impacto negativo ya que en comparación con las otras zonas geográficas esta fue la que menor probabilidad de pago obtuvo. La segunda variable fue la zona geográfica (65-80), dado que en la regresión logística se obtiene como resultado que las personas de esta zona tendrán menos probabilidad de pagar y en las redes neuronales sucede lo contrario. Al analizar las probabilidades (en base a la características iniciales que se mencionaron anteriormente en los apartados IV y V) se puede observar que donde se obtiene un mayor cambio es en el impacto negativo ya que estas disminuyen casi un 3%, en el impacto positivo el aumento que se tiene es menor al 1%. La siguiente variable es el giro manufactura la cual obtuvo un impacto negativo en la regresión logística aclarando que no es el giro que menor probabilidad tiene de pagar; en las redes neuronales su impacto fue positivo aunque el cambio fue mínimo. La variable que hace referencia a la experiencia que se haya tenido en otros negocios (antig) y la variable del tiempo que llevan con el negocio (actual) en la regresión logística ambas tuvieron un impacto negativo es decir, mientras menos experiencia tengan mayor será la probabilidad de pago, el resultado de las redes neuronales fue lo contrario mientras más experiencia tengan los microempresarios mayor será su probabilidad de pago, revisando ambas variables los cambios en la probabilidad fueron mínimos en las cuatro se tuvieron cambios del 0.20%, por lo tanto ninguna fue más significativa que la otra.

## VII. Conclusión

El *credit scoring*, actualmente es un tema muy novedoso, ya que en los últimos años han surgido más instituciones dedicadas a los otorgamientos de créditos, dando muchas facilidades para obtener estos. Las cuales necesitan de ciertas herramientas para tomar decisiones relacionadas con este tema. Actualmente el juicio humano ya no puede ser el único condicionante para las tomas de decisiones relacionado con el otorgamiento de créditos ya que es un método que incurre en muchos riesgos. Es por esto, que para reducir los riesgos y complementar el juicio humano, se implementan diversos modelos que hacen que las decisiones sean más precisas, ya que permiten un porcentaje de mayor acierto dado que con estos métodos se conocen más a fondo las características que tendría un buen pagador y un mal pagador, favoreciendo por un lado a las instituciones ya que tendrán mayores ganancias y de igual manera a las personas ya que en muchos casos se les negaba la posibilidad de obtener un crédito pensando que serían malos pagadores cuando en realidad era de manera contraria.

Los resultados de la mayoría de los modelos son muy similares, variando en los porcentajes de acierto y en qué tan complicado o no es de llevar a cabo al igual que en la manera de comprensión. De los diferentes modelos que existen para otorgar una calificación crediticia, dos de ellos son la regresión logística y las redes neuronales, siendo la primera uno de los métodos más utilizados y la segunda uno de los métodos más modernos. En la mayoría de las investigaciones mencionan a estos modelos como los mejores, siendo muy sencillos aplicarlos, entenderlos y con un gran porcentaje de confianza. En ambos se obtiene una probabilidad de pago dadas ciertas características de una persona, por lo cual resultó interesante realizar un trabajo donde estos dos modelos de *credit scoring* puedan ser analizados y comparados; y que mejor que aplicarlo a microempresarios mexicanos quienes son los más propensos a solicitar créditos para sus negocios, siendo este trabajo de gran ayuda para instituciones que se encuentran este país dedicadas al otorgamiento de créditos.



Para elaborar este material se utilizaron diversas características que pueden definir a un microempresario, pero sin duda alguna aún quedan más características que se pueden analizar para ampliar esta investigación y de esta manera mientras más características se tengan mayor será el porcentaje de acierto en la toma de decisiones.

Se sugiere continuar con el estudio de las características que hace que una persona cumpla con el pago de un crédito o no, ya que si bien en este trabajo se utilizaron diversas características, aún quedan varias por analizar. De igual manera se recomienda, investigar por qué los microempresarios mexicanos que cuentan con una carrera son los que menor probabilidad de pago tienen, cuando se piensa que las personas preparadas son las que mejor pagarían.

Por último se sugiere, continuar con la línea de estudios sobre las zonas geográficas en cuanto a porcentaje de pobreza, para analizar más a detalle por qué en los estados donde existe mayor pobreza es donde se tiene una mayor probabilidad de cumplimiento.

El *credit scoring* es un tema muy interesante y amplio por trabajar. En esta ocasión solo fue un estudio para microempresarios mexicanos, pero de igual forma se puede trabajar con otro tipo de empresas o instituciones siendo una herramienta muy útil para el tema de los créditos financieros.

## Bibliografía

- Ariza, M., Barón, W., Obregón, N., Pineda, Y., & Velosa, F. (2012). Modelo de selección de riesgo, aplicado al otorgamiento de microcrédito en un intermediario financiero de crédito colombiano, apoyado en redes neuronales artificiales. *Atlantic Review of Economics*, 1.
- Banda, H., & Garza, R. (2014). Aplicación teórica del método Holt Winterns al problema de credit scoring de las instituciones de microfinanzas. *Mercados y Negocios* 15(2), 5-21.
- BANXICO. (2005). Definiciones básicas de Riesgo. Recuperado el 10 de Marzo de 2016 de: <http://www.banxico.org.mx/sistema-financiero/material-educativo/intermedio/riesgos/%7BA5059B92-176D-0BB6-29587257E2799FAD%7D.pdf>
- BANXICO. (2013). Nuevos acuerdos de Basilea III. Recuperado el 20 de Marzo de 2016 de: <http://www.banxico.org.mx/sistema-financiero/material-educativo/basico/fichas/actividad-financiera/%7BF7AF1DAF-B3DE-4F00-DBCF-DC EE0E80C9E8%7D.pdf>
- Beltrán, M., Muñoz, A., & Muñoz, A. (2013). Redes bayesianas aplicadas a problemas de credit scoring. Una aplicación práctica. *Cuadernos de economía*, 37, 73-86.
- Bravo, C., Maldonado, S., & Weber, R. (2010). Experiencias prácticas en la medición de riesgo crediticio de microempresarios utilizando modelos de credit scoring. *Ingeniería de sistemas*. 24
- Calin, A., & Popovici, O. (Junio de 2014). Modeling credit risk throug credit scoring. *Internal Auditing anda Risk Managemet*, 2(34).
- Cardona, A. (2004). Aplicación de árboles de decisión en modelos de riesgo de crédito. *Revista Colombiana de Estadística*, 27(2), 139 - 151.
- Castromán, J.L. y Porto, N. (2000): "Asistencia Neuronal en la toma de n decisiones: una aplicación al credit scoring en un ambiente de

- incertidumbre”, VII Congreso de la Sociedad Internacional de Gestión y Economía Fuzzy (SIGEF), septiembre, Chania (Grecia).
- Cerezo, E., Claramunt, M., & Casanovas, M. (2011). Medición del riesgo de crédito mediante modelos estructurales: Una aplicación al mercado colombiano. *Cuadernos de administración*, 24(42), 73-100.
- Coloma, P., & Weber, R. (2006). Modelos analíticos para el manejo de riesgo de crédito. *Trend management*, 8.
- CONDUSEF. (2012). Microcréditos. Protege tu dinero. Recuperado el 20 de Mayo de 2016 de: <http://www.condusef.gob.mx/Revista/PDF-s/2012/152/micro.pdf>
- Dabos, M. (2010, Noviembre 2). Materia Biz. Recuperado el 30 de mayo de 2016 de: <http://www.materiabiz.com/mbz/economiafinanzas/nota.vsp?nid=4424>
- Escalona, A. (2011). Uso de los Modelos de Credit Scoring en Microfinanzas. Maestro en Ciencias. Institución de Enseñanza e investigación en Ciencias Agrícolas.
- Espin, O., & Rodríguez, C. (2013). Metodología para una scoring de clientes sin referencias crediticias. *Cuadernos Económicos*, 139-135.
- Esteve, E. M. (2007) Un modelo de credit scoring basado en el conocimiento de la aplicación de Basilea II y su papel innovador en el sector bancario. Recuperado el 10 de Febrero de 2016 de: <https://dialnet.unirioja.es/descarga/articulo/2499466.pdf>
- Gómez, L. (2008). Información Asimétrica: Selección Adversa y Riesgo moral. *Actualidad Empresarial* (3), 37-44
- Gómez, S. (2012). Revisión del uso de la lógica difusa aplicada a modelos de puntuación crediticia. *ACTIVA* 170.

- Gutiérrez, M. A. (2007). Modelos de credit scoring -Qué, cómo, cuándo y para qué- Recuperado el 15 de Febrero de 2016 de: <http://www.bcra.gov.ar/Pdfs/Publicaciones/CreditScoring.pdf>
- Hand, J., & Henley, E. (September de 1997). Statistical Classification Methods in Consumer Credit Scoring: a Review. *Journal of the Royal Statistical Society*, 160, 523-541.
- Ladino, C. (2014). Comparación de modelos de riesgo de crédito: Modelos logísticos y redes neuronales. Maestro en Economía. Pontificia Universidad Javeriana.
- Lara, J., Rayo, S., & Camino, D. (2010). Aspectos básicos del credit scoring y su implementación en las instituciones de microfinanzas en México. En V. Rodríguez, & D. Camino, Un acercamiento a la eficiencia del microfinanciamiento en México. Fundación Universidad de las Américas Puebla.
- Martínez, I. (2012). Definición y cuantificación de los riesgos financieros. *Tema de Portada* (30).
- Medina, E. (2003). Modelos de elección discreta. Recuperado el 23 de Septiembre de 2016 de: [http://www.uam.es/personal\\_pdi/economicas/eva/pdf/logit.pdf](http://www.uam.es/personal_pdi/economicas/eva/pdf/logit.pdf)
- Moreno, J. F., & Melo, L. F. (2011). Pronóstico de incumplimiento de pagos mediante máquinas de vectores de soporte: una aproximación inicial a la gestión del riesgo de crédito. *Borradores de economía*. 677
- Ochoa, J., Galeano, W., & Agudelo, L. (2010). Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera. *Perfil de coyuntura económica*, 191-222.
- Ortega, J., Martínez, J. y Valencia C. (2009). El modelo de calificación crediticia z-score. *MBA EAFIT*.

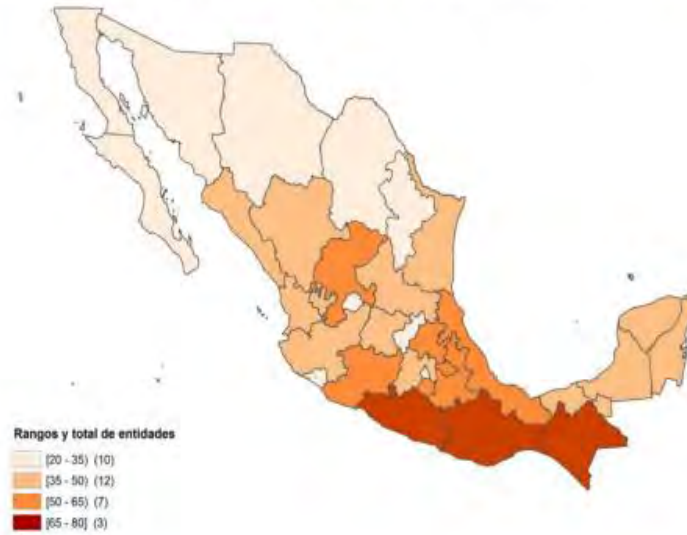
- Pagano, M., & Jappelli, P. (Diciembre de 1993). Information Sharing in Credit Markets. *Journal of Finance* , XLVIII (5).
- Pérez, F., & Fernández, H. (2007). Las redes neuronales y la evaluación del riesgo de crédito. *Ingenierías universidad de Medellín*, 6(10), 77-91.
- Porto, N., & Castromán, J. L. (2000). Nueva herramienta para la gestión: medición de la probabilidad de insolvencia, asistida por redes neuronales. Recuperado el 16 de Abril de 2016 de: <https://dialnet.unirioja.es/download/articulo/565243.pdf>
- Puertas, R., & Martí, M. L. (2012). Análisis del credit scoring. *RAE*, 303-315.
- Rojo J. (2007). Regresión con variable dependiente cualitativa. *Laboratorio de estadísticas*.1
- Rodríguez, V., Aguilar, P. (2013). Indicadores FOCO\* para medir la eficacia en la utilización del microcrédito. *FIR, FAEDPYME International Review*, 2(4), 73-88
- Rodríguez, V. (2015). “Línea Basal de Indicadores de micronegocios con y sin crédito en México”. IV Congreso Nacional CIMIPYME y I Congreso Internacional de Investigación sobre la Pequeña y Mediana Empresa de ICSB México. ICSB México, UNIVA, AIREPME México. Puerto Vallarta, Jalisco. Del 23 al 25 de septiembre de 2015.
- Rodríguez V., Hernández J. (2013) Matriz de probabilidad de transición de microcréditos: el caso de una microfinanciera mexicana. *Estudios Económicos*, 28 (1) pp. 39-77
- Schreiner, M. (1999). Un modelo de calificación del riesgo de morosidad para los créditos de una organización de microfinanzas en Bolivia. Recuperado el 19 de Abril de 2016 de: [http://www.microfinance.com/Castellano/Documentos/Bolivia\\_Scoring\\_Mora.pdf](http://www.microfinance.com/Castellano/Documentos/Bolivia_Scoring_Mora.pdf)
- SECQ. Modelo de Regresión logística (2008)

- Serrano, C., & Gallizo, J. L. (1996). Las redes neuronales artificiales en el tratamiento de la información financiera. Zaragoza. Recuperado el 16 de Abril de 2016 de: <http://ciberconta.unizar.es/Biblioteca/0004/SerGall96.html>
- Santos, C. (1999). Riesgo de Crédito en México. Citado en Riesgo de Crédito. Recuperado el 15 de Febrero de 2016 de: [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/laex/garcia\\_s\\_m/capitulo2.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/laex/garcia_s_m/capitulo2.pdf)
- Torrico, S. (Noviembre de 2014). Macro credit scoring como propuesta para cuantificar el riesgo de crédito. *Investigación y desarrollo*, 2(14), 42-64.
- Turner, M., & Robin Varghese. (27 de March de 2006). The Benefits of Wider Participation in Full-File Credit Reporting in Latin America and the Costs of the Status Quo. Information Policy Institute.

## Anexos

### Ilustración 1 Mapa de Clasificación

**Mapa 1. Porcentaje de la población en pobreza, según entidad federativa, 2014**



Mapa realizado por la CONEVAL, sobre la clasificación de los estados según el porcentaje de pobreza que tengan.

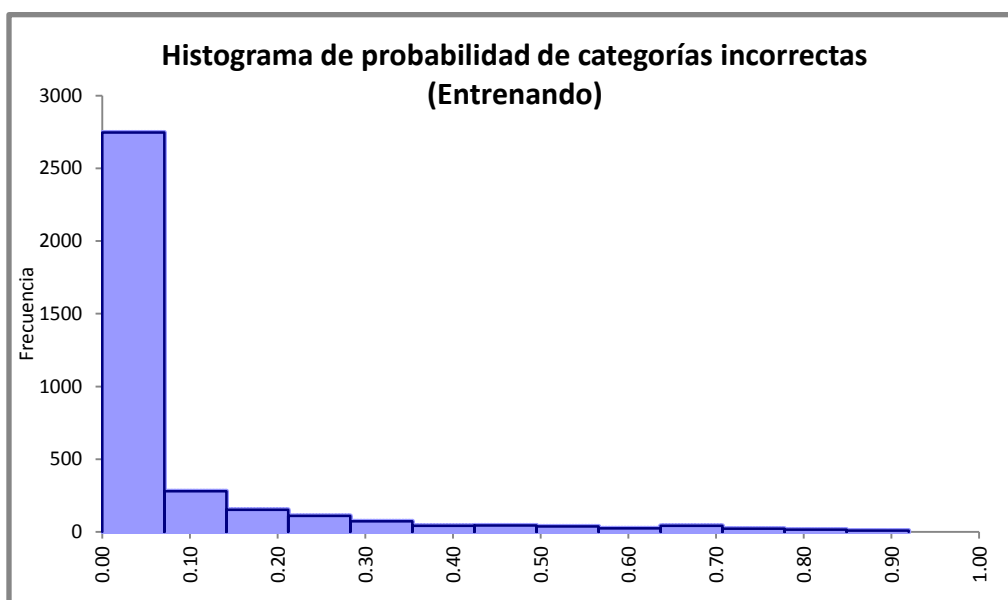
### Resumen de *neuraltools* \$4032

Resumen	
<i>Información de red</i>	
Nombre	Red entrenada en 4032
Configuración	Predicción de categoría PNN
Localización	Este libro de trabajo
Variable de categoría independiente	12 (sex, (20-35), (35-50), (50-65), (65-80), NIVEDU, MANU, COM, SERV, PRES_H, PRES_R, EDO_CIV)
Variables numéricas independientes	8 (EDAD, P, MENSP, MESES, VENTAS, GANAN, ANTIG, ACTUAL)
Variable dependiente	Var. de categoría (CRED4032)
<i>Entrenando</i>	
Número de casos	3590
Tiempo de Entrenamiento	0:03:59
Número de pruebas	88
Razón de la parada	Auto-Parada
% de predicciones incorrectas	4,0390%
Probabilidad incorrecta media	7,2516%
Desviación estándar de probabilidad incorrecta	15,4352%

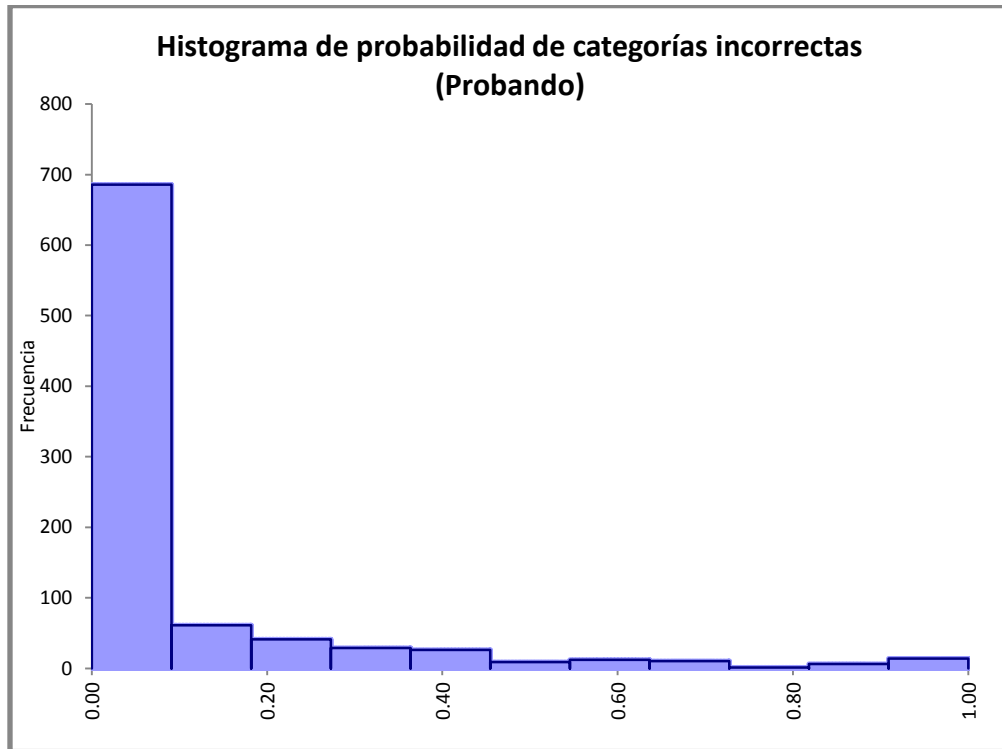
<b>Probando</b>	
Número de casos	898
% de predicciones incorrectas	5,6793%
Probabilidad incorrecta media	9,4161%
Desviación estándar de probabilidad incorrecta	19,8727%
<b>Predicción</b>	
Número de casos	1
Predicción en Vivo activada	SÍ
<b>Conjunto de datos</b>	
Nombre	4032
Número de filas	4489
Etiquetas manuales de caso	NO

<b>Matriz de clasificación</b>			
(para casos de entrenamiento)			
	0	1	Incorrecto (%)
0	1334	94	6,5826%
1	51	2111	2,3589%

<b>Matriz de clasificación</b>			
(para casos de prueba)			
	0	1	Incorrecto (%)
0	333	37	10,0000%
1	14	514	2,6515%







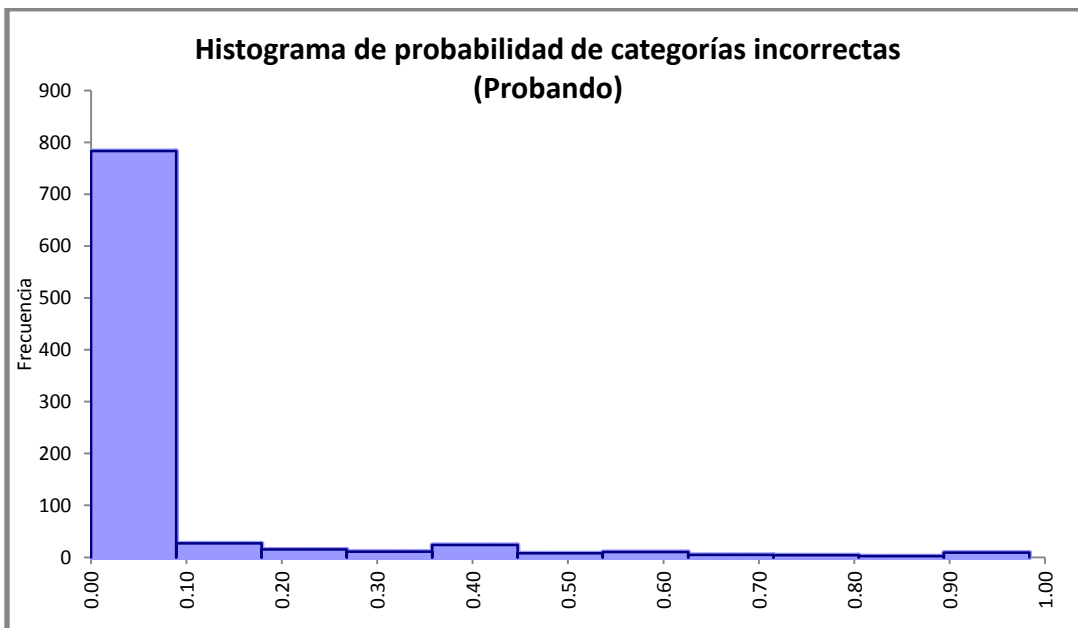
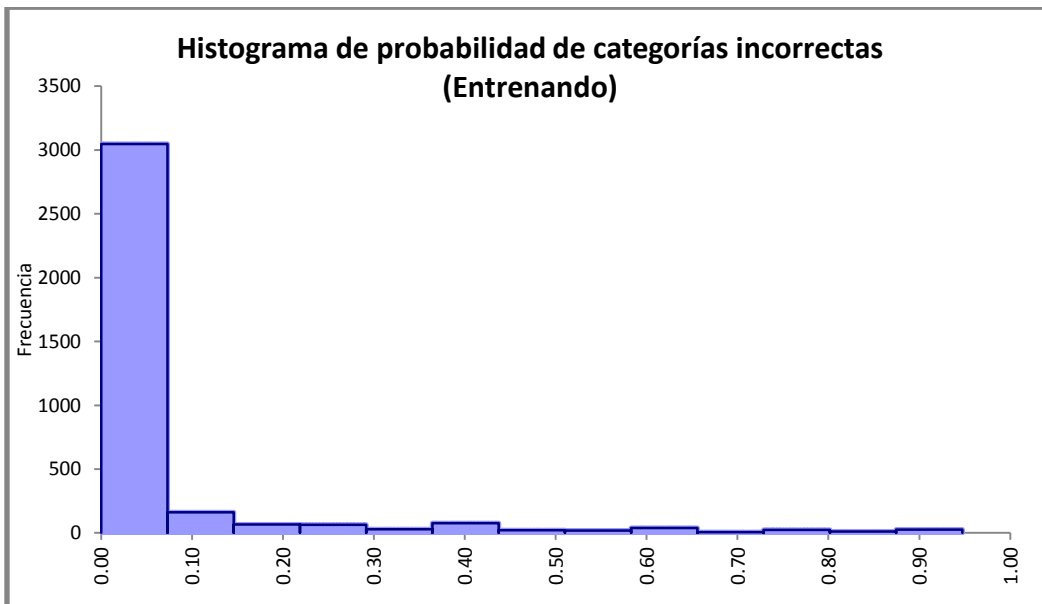
## Resumen de neuraltools \$8344

Resumen	
<i>Información de red</i>	
<b>Nombre</b>	Red entrenada en 8433
<b>Configuración</b>	Predicción de categoría PNN
<b>Localización</b>	Este libro de trabajo
<b>Variable de categoría independiente</b>	12 (sex, (20-35), (35-50), (50-65), (65-80), NIVEDU, MANU, COM, SERV, PRES_H, PRES_R, EDO_CIV)
<b>Variables numéricas independientes</b>	8 (EDAD, P, MENS, MESES, VENTAS, GANAN, ANTIG, ACTUAL)
<b>Variable dependiente</b>	Var. de categoría (CRED8344)
<i>Entrenando</i>	
<b>Número de casos</b>	3591
<b>Tiempo de Entrenamiento</b>	0:10:53
<b>Número de pruebas</b>	101
<b>Razón de la parada</b>	Auto-Parada
<b>% de predicciones incorrectas</b>	3,6202%
<b>Probabilidad incorrecta media</b>	5,7553%

<b>Desviación estándar de probabilidad incorrecta</b>	15,3222%
<b>Probando</b>	
<b>Número de casos</b>	898
<b>% de predicciones incorrectas</b>	3,4521%
<b>Probabilidad incorrecta media</b>	5,8340%
<b>Desviación estándar de probabilidad incorrecta</b>	15,7530%
<b>Predicción</b>	
<b>Número de casos</b>	1
<b>Predicción en Vivo activada</b>	Sí
<b>Conjunto de datos</b>	
<b>Nombre</b>	8433
<b>Número de filas</b>	4490
<b>Etiquetas manuales de caso</b>	NO

<b>Matriz de clasificación</b>			
(para casos de entrenamiento)			
	<b>0</b>	<b>1</b>	<b>Incorrecto (%)</b>
<b>0</b>	2042	104	4,8462%
<b>1</b>	26	1419	1,7993%

<b>Matriz de clasificación</b>			
(para casos de prueba)			
	<b>0</b>	<b>1</b>	<b>Incorrecto (%)</b>
<b>0</b>	510	21	3,9548%
<b>1</b>	10	357	2,7248%



**Tabla 16 Regresión comparando tipos de créditos (Refaccionario- Habilitación o avío) \$4,032**

Dependent Variable: CRED4032  
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)  
Date: 07/07/16 Time: 10:09  
Sample (adjusted): 1 4489  
Included observations: 4489 after adjustments  
Convergence achieved after 4 iterations  
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
PRES_H	0.002645	0.153409	0.017243	0.9862
PRES_R	0.557297	0.163368	3.411290	0.0006
C	0.274943	0.149243	1.842245	0.0654
<hr/>				
McFadden R-squared	0.009430	Mean dependent var	0.599465	
S.D. dependent var	0.490061	S.E. of regression	0.487125	
Akaike info criterion	1.335095	Sum squared resid	1064.488	
Schwarz criterion	1.339378	Log likelihood	-2993.621	
Hannan-Quinn criter.	1.336604	Deviance	5987.241	
Restr. deviance	6044.240	Restr. log likelihood	-3022.120	
LR statistic	56.99856	Avg. log likelihood	-0.666879	
Prob(LR statistic)	0.000000			
<hr/>				
Obs with Dep=0	1798	Total obs	4489	
Obs with Dep=1	2691			

**Tabla 17 Regresión comparando tipos de créditos (Refaccionario- Habilitación o avío) \$8,344**

Dependent Variable: CRED8344  
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)  
Date: 07/07/16 Time: 10:14  
Sample (adjusted): 1 4489  
Included observations: 4489 after adjustments  
Convergence achieved after 3 iterations  
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
PRES_H	-0.118987	0.155290	-0.766225	0.4435
PRES_R	0.424947	0.162853	2.609388	0.0091
C	-0.410021	0.150962	-2.716053	0.0066
<hr/>				
McFadden R-squared	0.009652	Mean dependent var	0.403653	
S.D. dependent var	0.490684	S.E. of regression	0.487548	
Akaike info criterion	1.337247	Sum squared resid	1066.338	
Schwarz criterion	1.341530	Log likelihood	-2998.450	
Hannan-Quinn criter.	1.338756	Deviance	5996.901	
Restr. deviance	6055.349	Restr. log likelihood	-3027.674	
LR statistic	58.44807	Avg. log likelihood	-0.667955	
Prob(LR statistic)	0.000000			
<hr/>				
Obs with Dep=0	2677	Total obs	4489	
Obs with Dep=1	1812			