

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

UNIDAD DE POST - GRADO

**Modelo de RNA para predecir la morosidad de
microcredito en la Banca Estatal Peruana**

TESIS

Para obtener el Grado Académico de Magíster en Ingeniería de Sistemas e
Informática con Mención en Ingeniería de Software

AUTOR

Luis Esteban Palacios Quichíz

Lima – Perú

2012

Modelo de RNA para predecir la morosidad de Microcrédito en la Banca Estatal Peruana

“Tesis presentada a la Universidad Nacional Mayor de San Marcos, (Lima – Perú), para optar el Grado de Magíster en Ingeniería de Sistemas e Informática, con la mención en Ingeniería de Software”.

Orientador: Mg. Fernando Torres Sánchez

UNMSM - LIMA

2012

FICHA CATALOGRÁFICA

17/09/2012

**Modelo de RNA para predecir la morosidad
de Microcrédito en la Banca Estatal Peruana**

Luis Esteban Palacios Quichíz

Lima – Perú, 2012

Orientador: Mg. Fernando Torres Sánchez

Disertación: Magíster en Ingeniería de Sistemas
e Informática.

Universidad Nacional Mayor de San Marcos

Escuela de Posgrado

Facultad de Ingeniería de Sistemas e
Informática, 2012

Unidad de Postgrado.

Páginas: 137.

DEDICATORIA

A la memoria de mis padres Luis Palacios Sánchez y Emma Quichíz de Palacios, porque siempre me inculcaron que el estudio es su mejor herencia.

A mi incansable esposa María Coronado Rodríguez por todo su apoyo brindado durante mis estudios de maestría.

Y a mis adorados hijos: Luigi, Ricardo, Antony y Esteban, para que les sirva de ejemplo en su vida futura.

AGRADECIMIENTOS

- A Dios, porque es la razón de nuestra existencia.
- Al profesor Luis Nieto, por su orientación y dedicación para que este trabajo cumpla con los objetivos trazados.
- A los profesores Dr. David Mauricio, Mg. Fernando Torres y Mg. Erick Vicente, por sus revisiones y mejoras del presente trabajo.
- A mi compañero de trabajo Wilbert Mamani que me ayudo en la programación del simulador convencional y en aprender juntos el uso de la Red Neuronal.
- A todas aquellas personas que indirectamente me ayudaron para culminar este trabajo.

RESUMEN

El Banco de la Nación tiene una delicada labor que cumplir ante los problemas sociales y económicos, para ampliar su cobertura a todo el territorio nacional y a todos los ciudadanos. Por ello conviene que **aumente radicalmente su competencia tecnológica** a fin de adoptar decisiones óptimas. Una de las formas tecnológicas de hacerlo es adoptar un modelo de **Red Neuronal Artificial (RNA)** que posteriormente será implementado y probado.

El presente trabajo de tesis pretende realizar una propuesta sobre un nuevo servicio de Microcrédito, el cual se intenta vincularlo a un aspecto muy puntual de la computación moderna aplicada, donde se muestre que es posible predecir la morosidad de los clientes, planteando un modelo basado en RNA. Este acercamiento innovador también incluirá la metodología de Minería de Datos para proyectos relacionados con redes neuronales artificiales.

PALABRAS CLAVES

Microcrédito, Redes Neuronales y Minería de datos.

ABSTRACT

The National Bank has a delicate work to play in social and economic problems, about extending their coverage throughout the national territory and for all citizens. Therefore it is important to increase radically their technological skills so that they know best decisions. One of the technological ways of doing so is adopting a model Artificial Neuronal network (RNA) which will then be implemented and tested.

The present thesis work made a proposal on a new service of microcredit, are trying to link it to a very specific aspect of applied modern computing, showing specifically that it is possible predict the delinquency customers, proposing a model based on RNA. This innovative approach also includes the Data Mining methodology for projects relationship with neural networks.

KEYWORDS

Microcredit, Artificial Neural Network and Data Mining.

INDICE

Tabla de figuras	xii
Tabla de cuadros	xiii
Capítulo I: INTRODUCCION	1
1.1. Antecedentes	1
1.2. Planteamiento del problema	3
1.3. Objetivos de la investigación	
1.3.1. Objetivo general	4
1.3.2. Objetivos específicos	4
1.4. Justificación	
1.4.1. Justificación teórica	4
1.4.2. Justificación práctica	5
1.5. Alcances	5
1.6. Propuesta de la solución	6
1.7. Organización de la tesis	7
CAPITULO II: MARCO TEÓRICO	9
2.1. Clientes de microcrédito	9
2.2. El Microcrédito en el Perú	12
2.3. Proceso del Microcrédito	15
2.4. Redes Neuronales Artificiales	
2.4.1. Elementos de una Red Neuronal Artificial	18
2.4.2. Perceptrón Multicapa	20
2.5. Método de Minería de Datos – CRISP - DM	22
2.6. Proceso de Modelamiento Unificado	25
CAPITULO III: ESTADO DEL ARTE	27
3.1. Clasificación de Microcréditos	27
3.2. Modelos Computacionales para Microcréditos	30
3.3. Justificación del uso de la Red Neuronal Artificial	46

CAPITULO IV: METODOLOGÍA DE LA INVESTIGACION	48
4.1. Selección de muestras	48
4.2. Plan de desarrollo del simulador convencional y de la RNA	50
4.3. Comparación de los artefactos	
4.3.1. Método de trabajo para la comparación	51
4.3.2. Técnicas de recojo de datos y su tratamiento	52
4.4. Determinación de requerimientos	
4.4.1. Hardware	59
4.4.2. Software	59
4.4.3. Sistemas operativos	60
4.4.4. Base de datos	60
CAPITULO V: INGENIERIA DE LOS ARTEFACTOS	61
5.1. Ingeniería del simulador convencional con RUP	
5.1.1. Captura de Requerimientos	61
5.1.2. Análisis y Diseño Orientado a Objetos	62
5.1.3. Implementación	62
5.1.4. Pruebas	63
5.2. Ingeniería de la Red Neuronal Artificial con CRISP-DM	
5.2.1. Comprensión del negocio	64
5.2.2. Comprensión de los datos	66
5.2.3. Preparación de los datos	68
5.2.4. Modelo de la red	70
5.2.5. Evaluación de la red neuronal	77
5.2.6. Despliegue o Implementación	77
CAPITULO VI: PRUEBAS – EXPERIMENTOS NUMERICOS	79
6.1. Primera corrida simulador convencional con la red neuronal	79
6.2. Experimentos numéricos	
6.2.1. Segunda corrida con la red entrenada	85
6.2.2. Tercera corrida con la red entrenada	89
6.2.3. Cuarta corrida con la red entrenada	93
6.2.4. Quinta corrida con la red entrenada	97
6.2.5. Resultados experimentos – Convencional versus RNA	101

CAPITULO VII: CONCLUSIONES Y RECOMENDACIONES	102
7.1. Conclusiones	102
7.2. Recomendaciones	102
REFERENCIAS BIBLIOGRAFICAS	104
ANEXOS	109
Anexo A1 Entrevista a experto en el servicio de microcrédito	109
Anexo A2 Entrevista a un Gerente Ejecutivo sobre el nuevo servicio y su importancia crítica para el banco	110
Anexo A3 Análisis documental de la conveniencia del servicio	112
Anexo A4 Análisis de la magnitud y restricciones del servicio	112
Anexo B1 Actores de prueba	114
Anexo B2 Diagrama de caso de uso de las pruebas	114
Anexo B3 Especificación de Caso de Uso: Simular Pago Convencional	115
Anexo B4 Diagrama de Secuencia	119
Anexo B5 Diagrama de clases diseño (con operaciones)	120
Anexo B6 Algoritmo – Programa Convencional – Montecarlo	121
Anexo B7 Construcción de la base de datos	123
Anexo B7.1. Diagrama de Clases (con atributos)	130
Anexo B7.2. Esquema de base de datos	130
Anexo B8 Diagrama de Componentes del proceso	133
Anexo B9 Diagrama de Despliegue	134
Anexo C Método de Montecarlo	135

Tabla de Figuras

Figura 1.1.	Esquema general del Trabajo de Investigación	8
Figura 2.1.	Proceso de Microcrédito	16
Figura 2.2.	Esquema básico de una Red Neuronal	18
Figura 2.3.	Esquema de una neurona artificial	19
Figura 2.4.	Perceptrón Multicapa	20
Figura 2.5.	Metodologías utilizadas en Data Mining	22
Figura 2.6.	Ciclo de Vida del Modelo CRISP-DM	23
Figura 3.1.	Dos ejemplos de GP usando operaciones matemáticas y condiciones	32
Figura 3.2.	Estructura de PNN	34
Figura 3.3.	Estructura de MLF Nets	35
Figura 3.4.	La estructura de VBDM	39
Figura 4.1.	Esquema de la obtención de la evidencia empírica	52
Figura 5.1.	Data de Prueba – Simulador Convencional	63
Figura 5.2.	Esquema del <i>Credit Scoring</i> con el RNA	64
Figura 5.3.	Data de Prueba - Red Neuronal	69
Figura 5.4.	Creación de la red neuronal	70
Figura 5.5.	Topología de la red Neuronal	71
Figura 5.6.	Configuración de controles para el entrenamiento	72
Figura 5.7.	Controles configurados	72
Figura 5.8.	Resultado del entrenamiento	73
Figura 5.9.	Curva de aprendizaje de la red	73
Figura 5.10.	Data de entrenamiento – Con cifra de mérito	74
Figura 5.11.	Grado de importancia de los factores objetivos	75
Figura 5.12.	Grado de sensibilidad de los factores objetivos	76
Figura 5.13.	Orden de importancia de los factores objetivos	76
Figura 5.14.	Resultado del entrenamiento	77
Figura 5.15.	Resultado de validación de prueba	77
Figura 5.16.	Data de prueba de validación – Sin cifra de mérito	78
Figura 6.1.	Resultado de los experimentos	101

Tabla de Cuadros

Cuadro 1.1. Evolución de Clientes en Microcréditos	1
Cuadro 1.2. Evolución de Microcréditos	2
Cuadro 2.1. Sistema Financiero Formal Peruano (SFFP)	17
Cuadro 2.2. Artefactos de RUP	25
Cuadro 3.1. Resultados – Comparación Programación Genética	33
Cuadro 3.2. Resultados – Comparación Red Neuronal con técnicas Convencionales	35
Cuadro 3.3. Resultados – Comparación Redes Neuronales, Programación Genética con técnicas convencionales - Bancos Privados	37
Cuadro 3.4. Resultados – Comparación Redes Neuronales, Programación Genética con técnicas convencionales - Bancos Públicos	38
Cuadro 3.5. Resultados – Comparación VBDTM con RSC	39
Cuadro 3.6. Resultados – Comparación VBDTM con otros modelos	40
Cuadro 3.7. Resultados – Comparación SVM con otros modelos	42
Cuadro 3.8. Resultados - Aprendizaje de Red Neuronal	43
Cuadro 3.9. Resultados - Aprendizaje de SVM	43
Cuadro 3.10. Resultados - Comparación Clasificador Híbrido	44
Cuadro 3.11. Resultados ACC – Técnicas Convencionales Vs. Técnicas Avanzadas	46
Cuadro 4.1. Datos Históricos Reales (6 meses anteriores)	53
Cuadro 4.2. Pronóstico programa Convencional (6 meses próx.)	53
Cuadro 4.3. Pronóstico programa RNA (6 meses próximos)	54
Cuadro 4.4. Correlación Convencional / Real	54
Cuadro 4.5. Correlación RNA / Real	55
Cuadro 4.6. Correlación RNA / Convencional	55
Cuadro 4.7. Diferencias por tipo y cantidad de omisiones Real / Convencional	57
Cuadro 4.8. T8: Diferencias por tipo y cantidad de omisiones Real / RNA	57
Cuadro 6.1.1. Datos Históricos Reales (6 meses anteriores)	79

Cuadro 6.1.2. Pronóstico del Programa Convencional (6 meses próximos)	80
Cuadro 6.1.3. Pronóstico del Programa RNA (6 meses próximos)	80
Cuadro 6.1.4. Correlación Convencional / Real	81
Cuadro 6.1.5. Correlación RNA / Real	81
Cuadro 6.1.6. Correlación RNA / Convencional	82
Cuadro 6.1.7. Diferencias por tipo y cantidad de omisiones Real / Convencional	83
Cuadro 6.1.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.	83
Cuadro 6.2.1.1. Datos Históricos Reales (6 meses anteriores)	85
Cuadro 6.2.1.2. Pronóstico del Programa Convencional (6 meses próximos)	85
Cuadro 6.2.1.3. Pronóstico del Programa RNA (6 meses próximos)	85
Cuadro 6.2.1.4. Correlación Convencional / Real	86
Cuadro 6.2.1.5. Correlación RNA / Real	86
Cuadro 6.2.1.6. Correlación RNA / Convencional	86
Cuadro 6.2.1.7. Diferencias por tipo y cantidad de omisiones Real / Convencional	86
Cuadro 6.2.1.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.	87
Cuadro 6.2.1.9. Datos de la segunda corrida	88
Cuadro 6.2.2.1. Datos Históricos Reales (6 meses anteriores)	89
Cuadro 6.2.2.2. Pronóstico del Programa Convencional (6 meses próximos)	89
Cuadro 6.2.2.3. Pronóstico del Programa RNA (6 meses próximos)	89
Cuadro 6.2.2.4. Correlación Convencional / Real	89
Cuadro 6.2.2.5. Correlación RNA / Real	90
Cuadro 6.2.2.6. Correlación RNA / Convencional	90
Cuadro 6.2.2.7. Diferencias por tipo y cantidad de omisiones Real / Convencional	90
Cuadro 6.2.2.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.	90
Cuadro 6.2.2.9. Datos de la tercera corrida	92

Cuadro 6.2.3.1. Datos Históricos Reales (6 meses anteriores)	93
Cuadro 6.2.3.2. Pronóstico del Programa Convencional (6 meses próximos)	93
Cuadro 6.2.3.3. Pronóstico del Programa RNA (6 meses próximos)	93
Cuadro 6.2.3.4. Correlación Convencional / Real	93
Cuadro 6.2.3.5. Correlación RNA / Real	94
Cuadro 6.2.3.6. Correlación RNA / Convencional	94
Cuadro 6.2.3.7. Diferencias por tipo y cantidad de omisiones Real / Convencional	94
Cuadro 6.2.3.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.	94
Cuadro 6.2.3.9. Datos de la cuarta corrida	96
Cuadro 6.2.4.1. Datos Históricos Reales (6 meses anteriores)	97
Cuadro 6.2.4.2. Pronóstico del Programa Convencional (6 meses próximos)	97
Cuadro 6.2.4.3. Pronóstico del Programa RNA (6 meses próximos)	97
Cuadro 6.2.4.4. Correlación Convencional / Real	97
Cuadro 6.2.4.5. Correlación RNA / Real	98
Cuadro 6.2.4.6. Correlación RNA / Convencional	98
Cuadro 6.2.4.7. Diferencias por tipo y cantidad de omisiones Real / Convencional	98
Cuadro 6.2.4.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.	98
Cuadro 6.2.4.9. Datos de quinta corrida	100

CAPÍTULO I

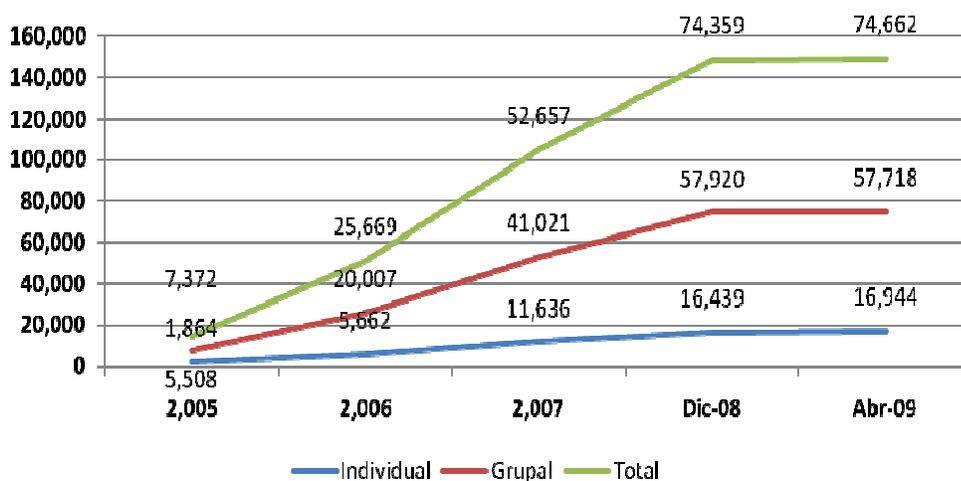
INTRODUCCIÓN

1.1. Antecedentes

El Banco de la Nación a partir del año 2006 emprendió un conjunto de reformas de modernización que la llevaron a ampliar la gama de servicios para sus clientes, multiplicar el número de sucursales, adoptar servicios electrónicos modernos. A pesar de lo dicho existe todavía **mucho espacio** para el mejoramiento tecnológico y el crecimiento de los servicios.

Uno de los servicios potenciales que muy bien podría ofrecer el banco es, el de **Microcrédito**, el cual ya ha sido asumido por diferentes entidades de crédito menores y con medios tecnológicos convencionales, mientras que los grandes bancos privados aún lo miran como difícil y por eso no se han animado a ofrecerlo todavía de manera clara a pesar de que saben muy bien que produce rendimientos interesantes, como se muestra en el cuadros siguientes, respecto a los créditos rurales: evolución de los clientes y a la evolución de los microcréditos.

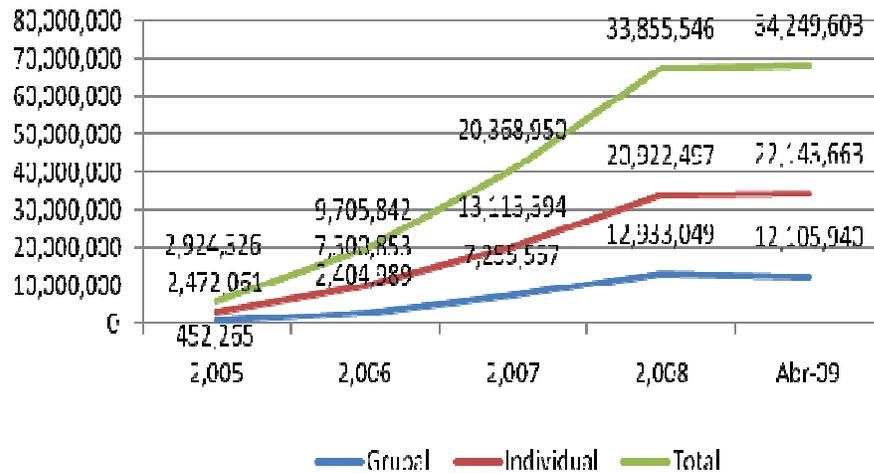
EVOLUCION CLIENTES



Cuadro No. 1.1. Evolución de Clientes en Microcrédito.

Fuente: Mi Banco (año 2009)

EVOLUCIÓN DE COLOCACIONES



Cuadro No. 1.2. Evolución de Microcréditos

Fuente: Mi Banco (año 2009)

Esta situación es pues **un problema y una oportunidad** simultánea para el banco y su solución conviene no solo al banco, sino a un gran segmento de su clientela potencial, aquella que es justamente la de más bajos ingresos. El **público objetivo** está formado por los clientes agrupados por compromisos de trabajo conjunto, mayormente mujeres, y de conducta de pago **muy poco predecible**, aunque la experiencia de las entidades que las sirven actualmente parece indicar que son en general serios en sus compromisos asumidos y que se esfuerzan por cumplirlo.

Otra razón para que el Banco de la Nación incursione, es que este servicio es de tipo social y por consiguiente contribuye con la inclusión social, el aumento del empleo y la lucha contra la pobreza por parte del Estado. En el año 2011, las cifras del Banco de la Nación revelan que ha realizado 130 mil operaciones de créditos a las micro financieras y canalizó S/. 530 millones en estas transacciones (Intranet del BN), eso demuestra que podría ingresar al mundo de los microcréditos para llegar al público objetivo.

Una tercera razón de mucho peso, es la necesidad para este nuevo Servicio de medios de control del repago más exactos y completos. El riesgo es elevado e impredecible con los métodos de predicción y cómputo acostumbrados en la banca.

Es necesario **innovar** en este campo y adoptar tecnologías de análisis de riesgo y de cómputo NO convencionales en el sector de la banca, aunque ya estén siendo usados en otros campos como en la industria eléctrica y las bolsas de valores. Se trata de la tecnología de computación neuronal o RNA que, a diferencia de la computación tradicional a base de algoritmos que descomponen los problemas en tareas formales específicas, lo que hace es **aprender de la experiencia** de manera similar al cerebro humano, tratando de averiguar **el perfil** del historial anterior del comportamiento del objeto (en este caso los clientes de microcrédito) para poder **predecir** su comportamiento futuro.

La ventaja de esta aproximación está en la **enorme sensibilidad a los cambios** en los perfiles de datos que obtiene este método de cómputo, en ciertos aspectos muy superior al humano y también al cómputo con algoritmos de la computación convencional. Por ello puede reemplazar con ventaja al mejor controlador humano de historiales de repago de clientes morosos.

Por tal motivo se plantea implementar un Modelo de Computación RNA para el servicio en cuestión a fin de evaluar en tiempo real los beneficios que cabrían esperar de su adopción por parte del banco.

1.2. Planteamiento del problema

El problema está en determinar un modelo apropiado, que permita predecir la morosidad de los clientes, para un eventual servicio de Microcrédito en el Banco de la Nación.

1.3. Objetivos

1.3.1. Objetivo general

El objetivo general para la presente investigación es: **Determinar un Modelo** apropiado para predecir la morosidad de los clientes en un futuro servicio de Microcrédito en el Banco de la Nación.

1.3.2. Objetivos específicos

Los objetivos específicos para la presente investigación son los siguientes:

- **Describir** la computación neuronal y su aplicación a la tarea de control de créditos de alto riesgo en el Banco de la Nación.
- **Establecer** los criterios de decisión concretos para elegir el software y/o hardware de cómputo neuronal óptimo para la tarea en cuestión.
- **Comparar** los rendimientos de la computación neuronal versus la computación convencional en la tarea de predecir el comportamiento de pagos futuros de clientes potenciales de microcrédito.
- **Desarrollar** un Modelo de RNA en el Banco de la Nación – en tiempo real y solo en primera aproximación – a fin de obtener todos los objetivos señalados.

1.4. Justificación

1.4.1. Justificación teórica

La contribución de este trabajo de investigación radica en el uso de modelos con redes Neuronales, para incursionar en diversos servicios financieros que puede otorgar la banca estatal, con la capacidad de predecir comportamientos **futuros de morosidad** en tiempo real de los clientes, con elevada eficiencia para otorgar el crédito o no. Actualmente, la Tasa de Morosidad en el Microcrédito es alta y los actuales modelos son deficientes. En Latinoamérica está entre 3.5 %, en el Perú entre 5 y 6 %.

Además de la aplicación de metodologías para la minería de datos, lo que apoya a modelar correctamente la data a ser cargada y entrenada con la red neuronal.

1.4.2. Justificación práctica

El beneficio es contribuir con la etapa de modernización del Banco de la Nación y la ampliación de su cobertura a todos los estratos sociales del Perú, con el diseño de nuevos servicios a sus clientes apoyando con la inclusión social y generando empleos en los sectores más pobres del país.

En la banca estatal se necesita predecir el comportamiento de los potenciales clientes según sus perfiles. Por lo tanto se requiere de controles (retrospectivos y prospectivos) precisos para predecir la morosidad, que se apoyen en la tecnología, para incursionar en el rubro del **Microcrédito**, el cual puede tener éxito si se liga desde su inicio con el **uso de las Redes Neuronales Artificiales**.

Las variables de la RNA a entrenar, son las más comunes que se utilizan actualmente en la realidad peruana, en un primer alcance; lo cual es flexible de poder agregarse nuevos factores objetivos de los clientes y poder retroalimentar (re-entrenar) a la RNA.

1.5. Alcances

Este trabajo muestra la pertinencia y adecuación del RNA como componente **esencial** para el diseño de nuevos servicios bancarios que demandan enfoques de computación no convencionales. El reto es establecer un Modelo computacional RNA como soporte del control del repago de los créditos de un servicio nuevo, en particular, el de determinar con exactitud la tasa de morosidad de todos sus clientes futuros en tiempo real.

La muestra seleccionada contiene datos de clientes del servicio de **Préstamos Multired** del Banco de la Nación, que se asemeja al futuro servicio de Microcrédito y sus préstamos solicitados son de montos pequeños al alcance de los sectores D y E de la población. Con esta muestra se calificó a los clientes con los factores objetivos utilizados para el entrenamiento de la RNA. Durante la determinación de las

variables de entrada (factores objetivos) se asumió que el cliente a evaluar ya tiene un historial crediticio (prestamos anteriores), en otras palabras es un cliente antiguo.

Esencialmente el modelo consta de una arquitectura simple compuesto por los siguientes módulos:

- Base de datos retrospectiva de clientes convencionales.
- Base de datos prospectiva de clientes de Microcrédito.
- Software de RNA y Software de Programación Convencional.
- Hardware convencional - una laptop.
- Programa de Entrenamiento de la red neuronal orientado a predecir el comportamiento crediticio de los clientes, tanto con la tecnología neuronal como con la convencional.
- Programa de comprobación estadística de **las diferencias significativas de predicción** entre ambas tecnologías y su validez objetiva respecto al nuevo servicio de Microcrédito.

1.6. Propuesta de la solución

Este trabajo no solo se limita a mostrar la computación neuronal y hacerlo construyendo un Modelo A con sus correspondientes arquitecturas de hardware y software, con componentes comerciales, sino también, en comparar sus capacidades funcionales frente a algún otro Modelo B y a partir de allí decir que el Modelo A es mejor que el Modelo B porque los valores de desempeño conseguidos son más altos (más velocidad, menos tiempo, etc.).

En este trabajo además se muestran las pruebas comparativas, entre el **cómputo neuronal** (reconocimiento de patrones) utilizando el paquete Easy NN Plus y el **convencional** (programa probabilístico) utilizando Java. Es decir se incluye un Modelo RNA riguroso de solución donde un módulo fundamental del mismo es la comprobación, estadísticamente fundamentada, de los datos comparativos que evidencian su superioridad sobre las soluciones convencionales en uso.

1.7. Organización de la tesis

En el capítulo I, se explica los antecedentes del banco respecto al tema de ampliar sus servicios y la oportunidad de incursionar en el servicio de microcrédito. Se explica el problema para determinar el tipo de software apropiado para el nuevo servicio. Se da una explicación básica de la propuesta de solución y, además, de la contribución que aporta al eventual nuevo servicio esta investigación.

En el capítulo II, se desarrolla el marco teórico, que incluye a todos los conocimientos esenciales relacionados al microcrédito., tecnología de información utilizada, las redes neuronales (perceptrón multicapa) y las metodologías CRISP-DM y RUP a utilizar.

En el capítulo III, se expone el estado del arte y, de manera muy concreta las experiencias que actualmente se tienen respecto a microcréditos con redes neuronales, que darán el soporte para la elaboración de la estrategia de solución.

En el capítulo IV, se explica la metodología a seguir para la implementación del Simulador Convencional y la Red Neuronal. Además de la selección de muestras y el procedimiento para la comparación entre ambas soluciones.

En el capítulo V, se ve la implementación de los artefactos, que parten desde el modelo conceptual hasta la explicación de una arquitectura de simulación convencional y la red neuronal artificial, las que concluyen con su despliegue respectivo.

En el capítulo VI, se efectúan las pruebas de entrenamiento de la red neuronal, midiendo el comportamiento y realizando los ajustes. Así, a partir de una serie de ensayos se comprobará el grado de eficiencia y aceptación de la red neuronal. Como parte esencial de la comprobación se utilizó el método estadístico de “**Diferencia de Medias de Pearson**”, sumando hasta cuatro (4) bloques experimentales de muestras

independientes de doscientos (200) clientes c/u (por la teoría de muestreo).

En el capítulo VII, se mencionan las conclusiones y recomendaciones, como resultado de la investigación.

La Figura 1.1. muestra la estrategia usada para la organización de la tesis, donde los capítulos II, III y IV se convierten en insumos para el capítulo V: Implementación de la solución. Luego se tienen las pruebas con experimentos numéricos y las conclusiones y recomendaciones en los capítulos VI y VII.

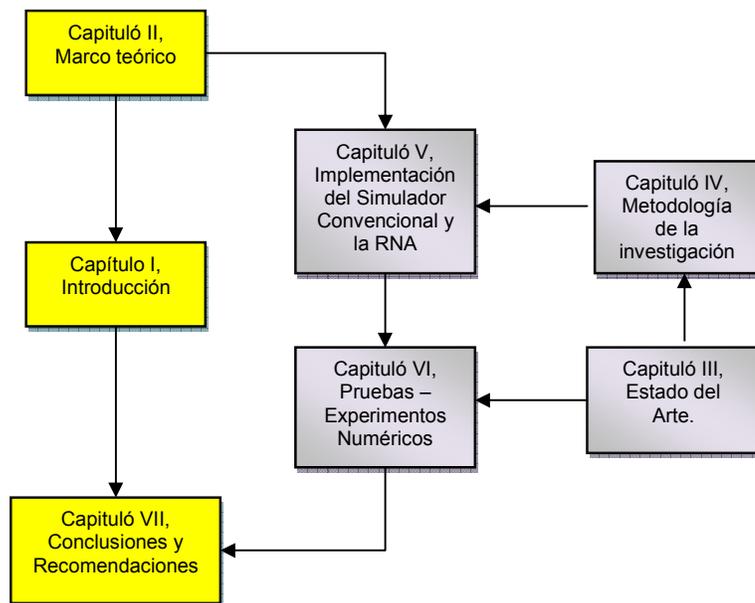


Figura No. 1.1. Esquema general del Trabajo de Investigación.

CAPÍTULO II

MARCO TEÓRICO

2.1. Clientes de microcrédito

Uno de los grandes retos que afrontan todos los países es la de conseguir la elevación de la **productividad** de las personas y empresas, ya que es esta una de las claves fundamentales sobre la que reposan los medios de acción para obtener las demás metas. La pobreza es una **expresión directa** de la insuficiente productividad de las personas y las sociedades, aunque naturalmente es solo uno de sus principales rasgos definitorios.

Puede definirse a la persona **pobre** como aquella cuya **productividad relativa** es lo suficientemente baja como para impedirle producir bienes o servicios vendibles en el mercado y suficientes para cubrir sus **necesidades básicas** de modo compatible con la dignidad humana (Dieterlen, 2003). Los **pobres extremos** (público objetivo potencial del nuevo servicio) son aquellos que suelen subsistir con alrededor de un dólar por día y normalmente se hallan excluidos “de facto” de la sociedad en su conjunto al no poder acceder al mercado ni hacer valer sus derechos humanos, que actualmente no tienen fuerza de ley. Caen en la mendicidad u otras actividades insuficientes y dañosas para sí mismos y para los demás.

Actualmente hay personas pobres, extremos o no, **en todos** los países del mundo (En los países desarrollados como los EUA el porcentaje es alrededor del 13% de toda la población, en los que se hallan en desarrollo como el Perú es del orden del 45% y en los atrasados como Bangladesh o Haití es del 75%). En general los estudios muestran que el 85% de la población mundial padece cierto grado de pobreza y un 35% padece pobreza extrema o crítica.

La economía de mercado pues no permite, **por principio**, eliminar a la pobreza pero sí es lo suficientemente flexible como para permitir

iniciativas heterodoxas que permiten mitigar en buena medida la pobreza **focalizada**, muy particularmente la de tipo extremo, **pero nada más**. Por ejemplo, en nuestro país hubo en los años 50 del pasado siglo un amplio Programa llamado Perú-Cornell auspiciado por la Universidad americana de ese último nombre para sacar a los campesinos andinos de la provincia de Vicos, Ancash, de la pobreza en base a sus propios esfuerzos organizados. El resultado fue bastante exitoso y logró demostrar que no hay **razones intrínsecas** atribuibles a los campesinos mismos que los obligue a vivir en la pobreza; siempre les era posible salir de ella cuando tenían el control de las variables básicas de sus comunidades. Infortunadamente, por razones de cambio de política, este programa fue cancelado, luego de algunos años, por 1970.

Sin embargo, recientemente ha surgido un mecanismo que despunta con gran vigor convirtiéndose en candidato, tal vez **el primero**, a ser de alcance mundial y efectivo a la vez: **La Banca Ética de Proyección Social** o Banca de los Pobres (Alsina, 2001). Esta es una de las más interesantes y prometedoras iniciativas heterodoxas que se han perfilado en los últimos años. **La Banca de los Pobres**, es la dinámica del **Microcrédito** institucionalizado para grupos organizados de pobres. Se originó en Bangladesh en los años 70, de allí se ha difundido por todo el mundo y se funda en tres Principios operativos sencillos pero altamente efectivos:

- La palabra del prestatario organizado
- El grupo de afinidad
- La intergarantía (garantía en grupo).

El Perú es uno de los países de Latinoamérica donde mejores resultados está dando este servicio de microfinanzas y todo indica que su futuro será brillante, sobre todo si logra salir de los escollos que lo entranpan actualmente. Algunos datos preliminares nos permiten apreciar el avance de este tipo de Banca en nuestro país:

- Hay un total de 584,187 MYPES¹ **formales** en 6 rubros económicos: Pesca, Minería, Manufactura, Construcción, Comercio, Otros Servicios, siendo el rubro de Comercio el de mayor número con 298,828 MYPES.
- Existe un total de 1,831,489 MYPES **informales** de los que 1,262,238 corresponden a Comercio
- El total de la PEA MYPES (Sin agro) es de 3,440,570 de los que corresponden 1,125,607 a Comercio
- Hay 2,4 millones MYPES y de ellas el 34% tiene saldos deudores con alguna entidad de crédito del Sistema Financiero²
- Hay 1,59 millones de MYPES **sin financiamiento** y se calcula que se podría financiar con el esquema clásico hasta un 24% con un promedio de US\$ 1,523 por MYPE. Queda pues **un buen margen** que debería ser apoyado con el nuevo concepto de Banca Ética y por ello es que afirmamos que este nuevo concepto tiene gran futuro en nuestro país y por ello merece mucha atención de los bancos nacionales y en particular del Banco de la Nación.
- Las tasas de interés anual de la Banca Múltiple a las MYPES es del orden de 26.68% mientras que a las grandes empresas solo les cobra 10.28% en promedio cuando son por más de 360 días. La razón que se aduce para tamaña diferencia (**supuesto** mayor riesgo) es simplemente insostenible y linda con la usura. Esta es una de las causas de las elevadas moras que se experimenta en el sector MYPE. Es irracional, y va contra los principios de la ciencia administrativa, que el que tiene menos pague más. Este absurdo es solo posible cuando prevalece la lógica del mercado libre, que es un automatismo ciego, y hay **ausencia del Estado** promotor, situación típica en nuestro país actualmente. Esto se podría revertir de raíz si entrara a tallar el Banco de la Nación que no solo realizaría una valiosa tarea de promoción social del pueblo peruano sino que de

¹ Debe diferenciarse entre MYPE (**Micro y Pequeña** Empresa) y PYME (**Pequeña y Mediana** Empresa) La primera conjuga mejor con la Banca Ética mientras que la segunda lo hace mejor con la Banca convencional.

² SALDAÑA NUÑEZ ALFREDO, **Op, Cit**, Pág. 99

paso obtendría una interesante utilidad que le permitiría ampliar el monto del capital dedicado a esta tarea.

Hay sin embargo un gran escollo en esta idea y es **el riesgo de mora** de parte de los prestatarios de este tipo de crédito que obliga a ser muy cuidadosos antes de incursionar en este terreno. Los errores típicos de las entidades de crédito que actualmente están dando este servicio son:

- Falta de **Planificación** estratégica (Visión, Misión, Objetivos, Metas, Análisis FODA)
- Errores en la **organización** de la entidad micro financiera
- Errores en los **procesos** de evaluación para el otorgamiento de los créditos
- Errores en la fase de **recuperación** de los micro créditos y
- Sensibilidad a factores exógenos del **entorno** socio-económico no previstos
- Sensibilidad a los eventos **aleatorios** en cualquier fase del proceso
- Falta de Control retroalimentado del tipo de **Auditoria Operacional** Integral Computarizada.

Un requisito fundamental que debe exhibir las soluciones candidatas para gestionar este nuevo servicio es la capacidad de **anticipar**, con 95% de confiabilidad mínima, el comportamiento de pagos del público objetivo, **mucho antes de que ocurran** estos pagos, examinando el patrón de sucesos asociados a la actividad del prestatario quien todavía ni siquiera sospecha que va camino al “default” es decir el incumplimiento de sus pagos. Allí está la fuerza del tipo de gestión que se necesita: la capacidad sobrehumana de reconocer patrones típicos aprendidos del comportamiento de los clientes por “experiencia” (aprendizaje) incesante.

2.2. El Microcrédito en el Perú

El Sistema Financiero Formal Peruano (SFFP) actualmente cuenta con 55 instituciones (48 dentro del segmento) reguladas por la SBS y está

formado por 15 Bancos, 13 Cajas municipales, 10 Cajas rurales, 13 Edpymes y 4 Financieras (Portal SBS, 2010).

- La primera institución microfinanciera de Perú fue creada en 1982 (CMAC Piura).
- Mi Banco es el primer banco privado, especializado en microfinanzas del Perú creado en 1998.
- Perú tiene uno de los más grandes mercados de microfinanzas en América Latina con aproximadamente 47 entidades reguladas participando en la industria.
- Actualmente, el sector de las microfinanzas en Perú es atendido principalmente por las siguientes instituciones: (i) 9 bancos, (ii) 13 CMACs; (iii) 10 CRACs; (iv) 12 Edpymes y (v) 3 Financiera.
- **CMACs:** Cajas Municipales de Ahorro y Crédito.
- **CRACs:** Cajas Rurales de Ahorro y Crédito.
- **Edpymes:** Entidades de Desarrollo de la Pequeña y Micro Empresa.
- **Empresa Financiera:** (Edyficar, TFC y Crediscotia).
- Total de instituciones reguladas por la SBS son 55. Ver cuadro 2.1.

De las experiencias del servicio de microcrédito resaltan los siguientes aspectos observados:

- a) Los créditos se dan a personas naturales y microempresarios
- b) Los clientes están localizados en áreas urbanas y rurales en todo el Perú.
- c) El crédito es ágil, sin una garantía formal.
- d) Inexistencia de garantías reales
- e) Existen contratos psicológicos
- f) El volumen del crédito va de acuerdo con el flujo de caja del cliente
- g) Una cuota vencida pasa a una situación transitoria, no a una morosidad total.
- h) Los asistentes de crédito hacen un reconocimiento de la situación, apreciación moral y renegociación al cliente.
- i) Existen clientes con baja sensibilidad a las tasas de intereses

- j) La atención es extremadamente personal y con entrenamiento al cliente.
- k) Existe un alto índice de renovación de cartera
- l) Énfasis en el análisis de crédito para la evaluación individual del cliente emprendedor
- m) Acompañamiento durante el desarrollo del proyecto del cliente
- n) El objetivo principal del crédito es el desarrollo y la inclusión social.
- o) Algunas entidades financieras permiten a sus clientes los accesos al sistema bancario y a créditos formales.

En el Perú es conocida la elevada correlación entre pobreza, el subempleo y la total carencia de empleo. En estas condiciones está claro que una reducción de los masivos niveles de pobreza, requiere como tarea previa elevar la productividad de la microempresa. Entre las políticas más destacadas para tal fin destaca la promoción del microcrédito.

El acceso al financiamiento, a igualdad de otros factores, puede incrementar la productividad, el volumen de producción y/o ventas, la inversión, fomentando así una mejora de los ingresos familiares y la creación de nuevos puestos de trabajo.

En esta tarea de clarificación de conceptos sobresale el trabajo de Felipe Portocarrero Maisch (Universidad del Pacífico, 2003), quien examina la oferta de microcrédito formal: Algunos bancos, ciertas ONG especializadas, cajas municipales, cajas rurales y Entidades de Desarrollo de la Pequeña y Microempresa (EDPYMES).

Ubica dicha oferta en un contexto del mercado financiero nacional, analiza las principales fuentes del microcrédito formal. Examina las características financieras, económicas e institucionales de las principales entidades de micro-financiamiento. Revisa el marco de regulación de estas instituciones, así como el papel que juega el Estado y la cooperación internacional en la promoción del sector.

Carolina Trivelli Ávila, actual ministra de Desarrollo e Inclusión Social, aborda el problema del microcrédito desde un ángulo poco explorado: el de los clientes. El documento presenta un balance de los estudios sobre la materia; examina las características generales de los usuarios de micro-finanzas y analiza los casos de tres instituciones que proveen microcrédito, accediendo a las respectivas bases de datos sobre sus clientes.

El acceso a las entidades formales resulta más restringido para los más pobres y aquellos con menor educación. En esas condiciones, los agentes informales logran una mayor presencia entre la población con menores recursos. La autora desarrolla un modelo que permite estimar la relación de diversas variables (edad, género, instrucción, lengua materna, tasa de dependencia y pobreza, entre otras) con la probabilidad de acceder al crédito.

Mientras se estaba procesando esta tesis, durante el año 2011, el BN empezó a dar pasos de gigante en este nuevo Servicio de Microcrédito, mucho más antes de lo que este autor creyó posible. Al cierre del 2011 el BN había logrado colocar nada menos que 652, 582 microcréditos a emprendedores de todo el país y establecer alianzas estratégicas con varias instituciones de intermediación financiera. El Banco dispone ahora de una línea de crédito de 200 millones de soles para canalizar a las MYPES vía las IFI. Ahora el Plan es colocar 799 millones de soles al cierre de este año 2012 (Revisa BN, Feb. 2012).

2.3. Proceso del microcrédito

El proceso crediticio en la microfinanzas cuando se quiere captar clientes nuevos pasa por las fases de: promoción (en una agencia o en el campo), evaluación, aprobación y desembolso. Posteriormente viene la fase de recuperación del crédito, según el periodo pactado con el cliente. Cuando un cliente ha tenido créditos anteriores, tiene un historial y ya no pasa por la fase de promoción (Salvador, y varios, 2010). Ver figura 2.1.

Para el caso de la promoción en el campo (sectores D – E para nuestro caso), el asesor de créditos toma los datos a los clientes y garantes a través de una declaración jurada, visitas diversas, verificaciones del negocio y a domicilio. Posteriormente en la agencia se hacen las revisiones de la evaluación y validaciones adicionales. Por último se hace la recomendación del crédito.

Los clientes realizan actividades de: cultivo de tierras, ganadería, avicultura, agroindustria, apicultura, comercio, servicios, producción y compra de equipos, herramientas y maquinarias; desarrolladas en el ámbito urbano (ambulante o emprendedor en un lugar establecido) y también el ámbito rural. Ellos ahorran mensualmente y pueden solicitar créditos en forma individual y grupal. En centros poblados o caseríos forman grupos solidarios:

- Grupo solidario tradicional (de 2 a 5 clientes), con un coordinador
- Grupo solidario ampliado (de 6 a 20 clientes), con una junta directiva y
- Bancos comunales.

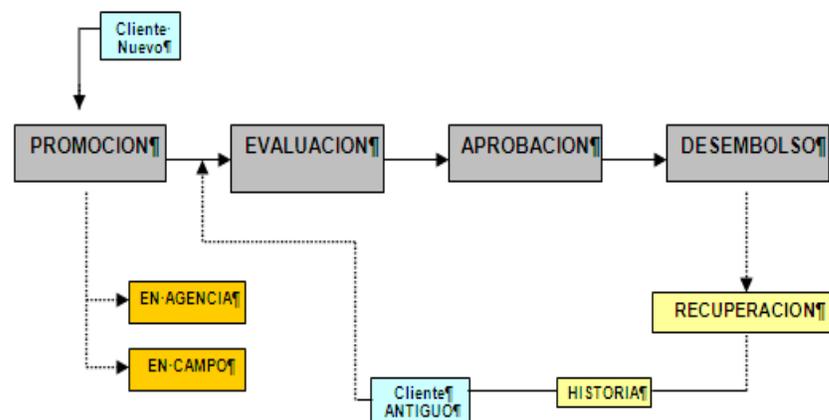


Figura No. 2.1. Proceso de Microcrédito

Fuente: Mi Banco (año 2009)

BANCA MULTIPLE	CMACs	CRACs	EDPYMES	EMP. FINANCIERAS
B. de Crédito del Perú	Arequipa	Nuestra Gente	Confianza	Crediscotía Financiera
B. Continental	Trujillo	Señor de Luren	Raiz	Solución Fina. de Crédito
Scotiabank Perú	Piura	Credinka	Crear Arequipa	Finaciera TFC S.A.
Interbank	Sullana	Profinanzas	Proempresa	Financiera Edyficar
Interamericano de Finanzas	Cusco	Chavín	Efectiva	
Mibanco	Huancayo	Cajamarca	Nueva Visión	
Citibank	Tacna	Sipán	Alternativa	
B. Financiero	Ica	Los Andes	Credivisión	
HSBC Bank Perú	Lima	Prymera	Micasita	
B. Falabella Perú	Paita	Libertadores de Ayacucho	Acceso Crediticio	
B. de Comercio	Maynas		Pro Negocios	
B. Ripley	Del Santa		Solidaridad	
B. Santander Perú	Pisco		Credijet	
B. Azteca Perú				
Deutsche Bank Perú				
15	13	10	13	4

Cuadro No. 2.1. Sistema Financiero Formal Peruano (SFFP)

Fuente: SBS (Superintendencia de Banca, Seguros y AFPs, 2010)

2.4. Redes Neuronales Artificiales

2.4.1. Elementos de una Red Neuronal Artificial

Las **redes de neuronas artificiales** (denominadas habitualmente en castellano como **RNA** (en inglés como **ANN**, por Artificial Neural Networks) pueden definirse en primera aproximación como redes interconectadas masivamente en paralelo de elementos básicos muy simples de tipo software o hardware con organización jerárquica, capaces de adoptar un comportamiento colectivo adaptativo con el que intentan interactuar con los objetos del mundo real de modo análogo como lo hace el sistema nervioso biológico (Blum, 1992).

Son modelos que intentan reproducir el comportamiento del cerebro. Los mismos constan de dispositivos elementales de proceso: *las neuronas*. A partir de ellas, se pueden generar representaciones específicas, de tal forma que un estado conjunto de ellas puede significar una letra, un número u otro objeto (Caudill, y varios, 1992).

En la figura 2.2. se compara una neurona biológica con una neurona artificial. En la misma se pueden observar las similitudes y diferencias entre ambas (tienen entradas, utilizan pesos y generan salidas).

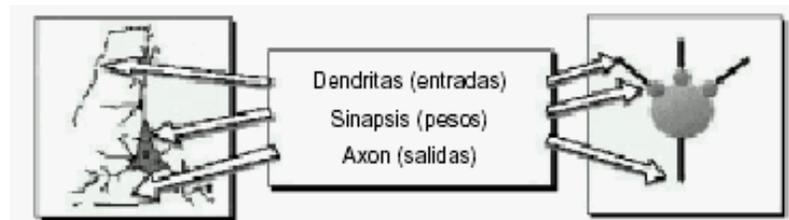


Figura 2.2. Neurona Biológica y Artificial
(Caudill, y otros, 1992)

La neurona artificial pretende mimetizar las características más importantes de las neuronas biológicas (Hertz, y varios, 1991). Cada neurona i -ésima está caracterizada en cualquier instante por un valor numérico denominado valor o estado de activación $a_i(t)$; asociado a cada unidad, existe una función de salida, f_i que transforma el estado actual de activación en una señal de salida. Dicha señal es enviada a

través de los canales de comunicación unidireccionales a otras unidades de la red; estos canales permiten que la señal se modifique de acuerdo con la sinapsis (el peso, w_{ji}) asociada a cada uno de ellos según determinada regla. Las señales moduladas que han llegado a la unidad j -ésima se combinan entre ellas, generando así la entrada total Net_j . Cuyo valor se define con la función:

$$Net_j = \sum_i y_i w_{ji}$$

Una función de activación, F , determina el nuevo estado de activación $a_j(t+1)$ de la neurona, teniendo en cuenta la entrada total calculada y el anterior estado de activación $a_j(t)$. El esquema de la función se aprecia en la figura 2.3.

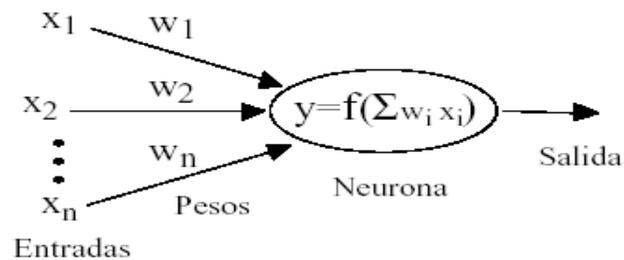


Figura 2.3. Esquema de una Neurona Artificial
(Hertz John, y otros, 1991)

La dinámica que rige la actualización de los estados de las unidades puede ser de dos tipos: asíncrono y síncrono. En el primer caso, las neuronas evalúan su estado continuamente según les va llegando información, y lo hacen de forma independiente, En el segundo caso, la información llega de forma continua, pero los cambios se realizan simultáneamente, como si existiera un reloj interno que decidiera cuando cambiar su estado. Los sistemas biológicos se hallan probablemente entre ambas posibilidades. Se han usado particularmente las referencias (Isasi, y varios, 2004), (Sánchez, y varios, 2006), (Hilera, y varios, 2000) y (Del Brio y Sanz, 2002) de la bibliografía propuesta.

2.4.2. El Perceptrón Multicapa

Es una red neuronal artificial formada por múltiples capas de neuronas, agregadas a fin de superar la linealidad del perceptrón simple y resolver los problemas que no son linealmente separables, lo cual es su principal limitación.

El perceptrón multicapa puede estar total o localmente conectada. En el primer caso cada salida de una neurona de la capa "i" es entrada de todas las neuronas de la capa "i+1", mientras que en el segundo cada neurona de la capa "i" es entrada de una serie de neuronas (región) de la capa "i+1". Ver figura 2.4.

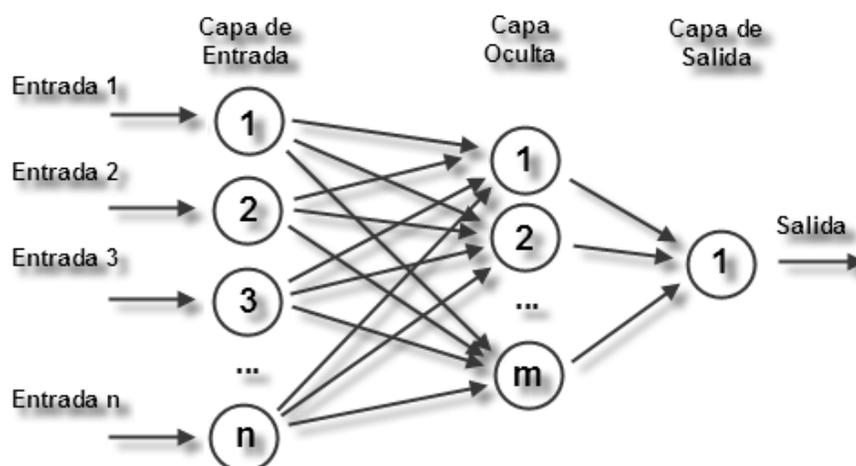


Figura 2.4. Perceptrón Multicapa
(Bernacki Mariusz, y otros, 2004)

Las capas se clasifican en tres tipos:

- **Capa de entrada:** Constituida por aquellas neuronas que introducen los patrones de entrada en la red. En estas neuronas no se produce procesamiento.
- **Capas ocultas:** Formada por aquellas neuronas cuyas entradas provienen de capas anteriores y cuyas salidas pasan a neuronas de capas posteriores.
- **Capa de salida:** Neuronas cuyos valores de salida se corresponden con las salidas de toda la red.

La propagación hacia atrás (también conocido como retro - propagación del error o regla delta generalizada), es un algoritmo utilizado en el

entrenamiento de estas redes, ya que proporciona la reactualización de los pesos usados por las neuronas para computar su disparo, por ello, el perceptrón multicapa también es conocido como red de retropropagación (no confundir con la red de contra - propagación).

En 1969, Minsky y Papert, demuestran que el perceptrón simple y ADALINE no puede resolver problemas no lineales (por ejemplo, XOR). La combinación de varios perceptrones simples podría resolver ciertos problemas no lineales pero entonces no existía un mecanismo automático para adaptar los pesos de la capa oculta. Rumelhart y otro autores, en 1986, presentan la "Regla Delta Generalizada" para adaptar los pesos propagando los errores hacia atrás, es decir, propagar los errores hacia las capas ocultas inferiores. De esta forma se consigue trabajar con múltiples capas y con funciones de activación no lineales.

Se demuestra que el perceptrón multicapa es un aproximado universal. Un perceptrón multicapa puede aproximar relaciones no lineales entre los datos de entrada y salida. Esta red se ha convertido en una de las arquitecturas más utilizadas en el momento (Bernacki, y otros, 2004). El perceptrón multicapa (de aquí en adelante MLP, MultiLayer Perceptron) se utiliza para resolver problemas de asociación de patrones, segmentación de imágenes, compresión de datos, etc.

Compresión de datos

Considérese un MLP de 3 capas, una de entrada, una oculta y la de salida. La capa de entrada está formada por N neuronas, la capa oculta por M ($M < N$) neuronas y la capa de salida posee N neuronas al igual que la capa de entrada. Se entrena dicho MLP para que cuando se le dé como entrada un vector de datos (x_1, x_2, \dots, x_N) devuelva ese mismo vector con M datos como salida, con ello estamos enseñando al MLP a transformar un vector de N componentes en uno de M componentes (recordemos que $M < N$) y a recuperar el vector original a partir del vector "comprimido".

Una vez que el MLP esté entrenado se procede de la siguiente forma:

- **Compresión:** Para comprimir los datos se utiliza un MLP de dos capas, la de entrada con N neuronas y la de salida con M, los pesos de estas dos capas son los de la capa de entrada y oculta respectivamente, del MLP que entrenamos anteriormente.
- **Descompresión:** Para descomprimir los datos se utiliza un MLP de dos capas, la de entrada con M neuronas y la de salida con N, los pesos de estas dos capas son los de la capa oculta y la de salida respectivamente, del MLP que entrenamos anteriormente. El MLP no conseguirá (al menos normalmente) un error nulo durante el entrenamiento, por lo que se trata de un sistema de compresión con pérdidas. Obviamente cuanto mayor se quiere que sea el factor de compresión, más errores aparecerán.

2.5. Método de minería de datos – CRISP- DM

Los actuales modelos de procesos de minería de datos (Chapman, y otros, 2000) en sus marcos de trabajo se sustentan haciendo uso de un ciclo de vida, las mismas que van modelando el comportamiento de los datos y de los conocimientos, y así obtener un resultado que tenga significado en el mundo real (Valcárcel Ascencios, 2011). Según la figura 2.5. CRISP-DM (*Cross – Industry Standard Process for Data Mining*) es una de las metodologías más usadas.

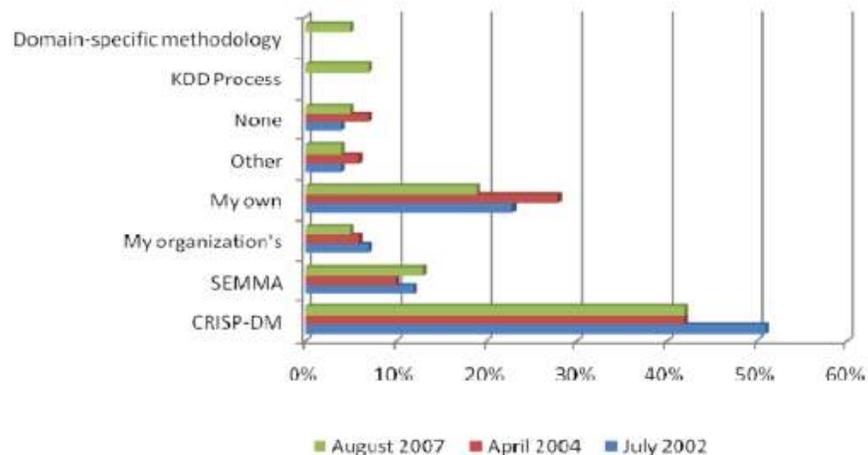


Figura 2.5. Metodologías utilizadas en Data Mining
(Gallardo, 2009)

La metodología CRISP-DM estructura el ciclo de vida de un proyecto de explotación de datos en seis fases cuya sucesión no es rígida e interactúan entre ellas de forma iterativa durante el desarrollo del proyecto. Las flechas indican la dependencia más importante y frecuente entre las fases. El círculo exterior simboliza la naturaleza cíclica de los proyectos de este tipo. Ver figura 2.6.



Figura 2.6. Ciclo de Vida del Modelo CRISP-DM
(Chapman, y otros, 2000)

A continuación se explican las fases en que se divide CRIP-DM:

- **Fase 1.** Comprensión del negocio o problema
Se centra en la comprensión de los objetivos y requisitos del proyecto desde una perspectiva de la empresa, para convertirlos en objetivos técnicos y en un plan de proyecto. El conocimiento adquirido del negocio se convierte en un problema de Data Mining (DM).
Las tareas que se realizan son: determinar los objetivos del negocio, evaluar la situación, determinar los objetivos del DM y elaborar un plan de proyecto.
- **Fase 2.** Comprensión de los datos
Comienza con una recopilación inicial de datos para establecer un primer contacto con el problema, familiarizarse con los datos, identificar la calidad de los datos y establecer relaciones evidentes

que permitan definir la hipótesis de información oculta.

Las tareas que se realizan son: recolectar los datos iniciales, describir los datos, explorar los datos y validar la calidad.

- **Fase 3.** Preparación de los datos

Competen las actividades de preparación de los datos iniciales para adaptarlos a las técnicas de DM, tales como: visualización, búsqueda de relaciones entre variables para la exploración.

Las tareas que se realizan son: selección de datos, limpieza de datos, estructurar los datos, integra los datos y formatear los datos.

- **Fase 4.** Modelado de datos

Diversas técnicas de modelado son seleccionadas y aplicadas y sus parámetros son calibrados a valores óptimos, por ejemplo análisis de regresión, redes neuronales, razonamiento basado en casos (RBC), etc. Algunas técnicas tienen requerimientos específicos según las características de los datos y la precisión que se requiera.

Las tareas que se realizan son: selección de la técnica de modelado, generar el plan de pruebas, construir el modelo y evaluar el modelo.

- **Fase 5.** Evaluación del modelo

Una evaluación detallada del modelo y la revisión de los pasos ejecutados para construir el modelo para asegurar que se han alcanzado los objetivos de negocio. Un punto importante es determinar si hay algún objetivo que no ha sido considerado.

Las tareas que se realizan son: evaluar los resultados, revisar el proceso y determinar la implementación.

- **Fase 6.** Despliegue o implementación

El conocimiento ó resultado adquirido debe ser documentada y presentada de manera comprensible para que el cliente pueda usarla. Sin embargo, dependiendo de la cantidad de requerimientos, está fase puede ser sencilla o compleja.

Las tareas que se realizan son: elaborar un plan de implantación,

elaborar un plan de monitoreo y mantenimiento, informar sobre el proyecto y la revisión del proyecto.

El método CRISP-DM es el más usado en el desarrollo de la minería de datos (Cobos y otros, 2010), y se aplica como guía metodológica para el desarrollo de las redes neuronales artificiales en esta investigación. En nuestro estudio utilizaremos esta metodología, para modelar la RNA, entrenarla con la información de los clientes de microcrédito y sus factores objetivos, validar su aprendizaje hasta su implementación.

2.6. Proceso de Modelamiento Unificado - RUP

Durante los últimos años, una de las metodologías más populares para el ciclo de vida de desarrollo de un software, ha sido el RUP (Rational Unified Process) de (Kruchten, 2000), es un proceso de ingeniería de software que ofrece un enfoque disciplinado para asignar tareas y responsabilidades a los roles participantes. Permite modelar el software a desarrollar visualmente usando el Unified Modeling Language (UML).

Esta guía metodológica cubre seis (6) flujos de trabajo principales: modelado del negocio, captura de requisitos, análisis y diseño, implementación, pruebas y distribución. Además de tres (3) flujos de soporte: administración de configuración y cambios, administración del proyecto y administración del entorno.

Para el presente trabajo de investigación sólo se aplican las disciplinas de captura de requerimientos, análisis y diseño, implementación y pruebas; para crear algunos artefactos del simulador convencional. Los artefactos a crear se muestran en el cuadro siguiente:

Captura de Requisitos	Diseño Orientado a objetos
1. Actores	1. Diagrama de secuencia
2. Casos de uso	2. Diagrama de clases de diseño
3. Diagrama de casos de Uso	3. Diagrama de tablas (entidades)
4. Especificación de caso de Uso	4. Diagrama de componentes
	5. Diagrama de despliegue

Cuadro No. 2.2. Artefactos de RUP (Fuente: Autor)

Cabe resaltar que los artefactos que vamos a crear son para darnos un mayor entendimiento de los casos de uso de prueba a desarrollar. Además, para especificar claramente los casos de uso (Cockburn, 2000). Se diseñaran todas las clases identificadas en la captura de requerimientos y se diseñará el comportamiento en su respectivo diagrama de secuencia (Braude, 2003). Finalmente se obtendrán las entidades del sistema (futuras tablas de base de datos).

CAPÍTULO III

ESTADO DEL ARTE

Este capítulo tiene el propósito de divulgar diferentes temas de investigación que tienen relación con la propuesta antes descrita, haciendo notar en qué medida se puede colaborar en lo que respecta a los modelos de cómputo orientados al microcrédito. En nuestro medio no es muy frecuente la aparición de trabajos con esta orientación, sin embargo el auge que está cobrando el tema de la Calificación del Crédito (Credit Scoring) en todo el mundo es de extremada competencia de nuevos servicios financieros y del uso de nuevas tecnologías.

3.1. Clasificación de Microcréditos **(Salvador Rayo, y otros, 2010)**

El crecimiento de los microcréditos a nivel mundial, junto con la normativa internacional sobre requerimientos al capital (Basilea II), están impulsando a las instituciones de micro-finanzas (IMF) a una mayor competencia entre las entidades financieras por este segmento de negocio. La banca tradicionalmente ha contado con adecuados modelos de Calificación de crédito para analizar el riesgo de incumplimiento. Sin embargo esto no ha sido así en las IMF supervisadas.

El objetivo de esta investigación es diseñar un modelo de calificación de crédito para una institución sometida a supervisión y especializada en microcréditos, como es la Entidad de Desarrollo de la Pequeña y Micro Empresa (Edpyme) del sistema financiero del Perú. El resultado de la investigación muestra la metodología y las fases necesarias para diseñar el modelo, así como el proceso de valoración y validación para que pueda ser aplicado en el área de negocio, especialmente para establecer la política de tasas de interés con clientes. Por último, también se muestra cómo puede utilizarse el modelo para desarrollar una gestión del riesgo de crédito en el marco de los métodos basados en calificaciones internas (IRB) de Basilea II.

(Yang Liu, 2002)

Con el desarrollo continuo y el cambio en la industria de crédito, los productos de crédito desempeñan un papel cada vez más importante en la economía. La globalización económica y los canales emergentes de servicio como Internet ofrecen posibilidades para que los clientes busquen y elijan a sus acreedores sin limitaciones regionales y el tiempo. Debido a esta tendencia, el acreedor debe ahora estar listo, dispuesto y capaz de extender el crédito a las empresas en otros países alrededor del mundo. El otorgamiento de crédito adopta pues una competencia mundial más drástica. El aumento de la demanda y el aumento de la competencia resultante de la nueva coyuntura económica ofrecen nuevas oportunidades, pero también presenta nuevas exigencias a las instituciones de la concesión de crédito.

Ellos buscaban con urgencia un ahorro de costes y una mayor eficiencia. Esto llevó a las instituciones a invertir el papel de la tecnología en su proceso de gestión de crédito.

Conforme aumentaba el volumen de créditos, el volumen de créditos insolventes presentaba una tendencia creciente también. En Alemania, el número de insolvencias presentadas por las empresas en el año 2000 aumentó un 6,6% desde 26,476 (el número en 1999) a 28,235, y en 2001 el número continuó aumentando a un ritmo aún mayor hasta un 14,3% a 32,278. El número de insolvencias de los consumidores era relativamente bajo en 1999 debido al impase legal que obligaba a la liquidación como condición previa a la petición de bancarrota.

En el año 2000, sin embargo, ese número se triplicó a un total de aproximadamente 10,500 casos. En el primer semestre de 2001, el número subió en un 50% a 6,807 casos. Las instituciones financieras y las instituciones de regulación prestaron más atención al mayor riesgo asociado al crédito. Las instituciones financieras invirtieron en recursos considerables para desarrollar herramientas eficientes y sofisticadas para evaluar y controlar los riesgos de crédito.

La Calificación de crédito es un modelo tecnológico que puede suministrar la parte básica de un sistema de apoyo a la toma de decisiones efectivas de crédito para atender las nuevas necesidades. Los modelos de Calificación de crédito implican las técnicas que se llaman hoy en día técnicas de minería de datos.

Los métodos de clasificación son las técnicas más utilizadas de minería de datos que se aplican en este ámbito para predecir el nivel de riesgo de las instituciones de crédito. Hay métodos de aprendizaje de máquina, y métodos estadísticos tales como la regresión lineal y logística, programación lineal, árboles de decisión, redes neuronales, etc. utilizados para el desarrollo de sistemas de Calificación de crédito.

En este trabajo se introduce, en el Capítulo 1, el conocimiento básico de la Calificación de crédito. Sus orígenes, el desarrollo y los problemas actuales son revisados. El énfasis no es en las técnicas de Calificación sino en los problemas de aplicación. El capítulo 2 se refiere a la definición de la Calificación de crédito. En el capítulo 3 se indica cómo enfrentar al nuevo entorno económico, conforme las áreas de aplicación de métodos de Calificación se expanden. La promoción de la aplicación de la Calificación de crédito y sus limitaciones también se discuten.

(Dean Karlan and Jonathan Zinman, 2007)

Un ingrediente clave de las estrategias de desarrollo en todo el mundo para ampliar el acceso al crédito de parte de operadores de micro-finanzas, políticos y donantes con ideas ambiciosas dedicados a la ampliación de este acceso, es buscar nuevos métodos eficaces para la aplicación y evaluación de esta expansión. Por ejemplo métodos de acceso al azar.

Hay menos consenso sobre el papel del crédito al consumo en las iniciativas de expansión. Algunas instituciones de micro-finanzas se están moviendo más allá del crédito empresarial y desean ofrecer créditos de consumo. Sin embargo, muchos profesionales y

responsables políticos se muestran escépticos acerca de lo "improductivo" de tales préstamos.

Estas preocupaciones son reforzadas por el trabajo académico que destacan las tendencias de comportamiento que pueden inducir a los consumidores a endeudarse más. Estiman que el impacto de una expansión de la oferta de crédito a los consumidores puede hacerse mediante un experimento de campo y el seguimiento paralelo con la obtención de datos. Un prestamista de Sudáfrica puede relajar sus criterios de evaluación de riesgos instruyendo a sus empleados de crédito que aprueben solicitudes seleccionadas al azar que fueron rechazadas. Se calcula entonces los impactos resultantes con los nuevos datos de la encuesta en los hogares de los candidatos a la adhesión y los datos administrativos sobre el reembolso de los préstamos, así como los informes de crédito público de uno y dos años más tarde.

Ellos encontraron que los préstamos marginales produjeron beneficios significativos para los prestatarios a través de una amplia gama de resultados económicos y de bienestar. También encontraron algunas evidencias de que los préstamos marginales eran rentables para el prestamista. Los resultados sugieren que la expansión del microcrédito al consumo puede ayudar bastante a la mejora del bienestar de los solicitantes.

3.2. Modelos Computacionales para Microcréditos

Models using Soft computer Methods.

(Adel Lahsasna, y otros, 2010)

En los últimos quince años, han atraído mucha curiosidad la aplicación de técnicas "blandas" de computación en el modelado de sistemas de control de créditos. Muchos modelos se han desarrollado utilizando métodos como las redes neuronales, los algoritmos genéticos o los métodos híbridos como la neuronal difusa o la genética difusa. Algunos de estos métodos se han aplicado y sus principales ventajas y

desventajas se han discutido.

Las redes neuronales, los algoritmos genéticos y las máquinas de soporte vectoriales han probado y reportado su valor claramente mayor que la de la mayor parte de los usuales métodos precisos. Por ejemplo la precisión de la clasificación de los clientes es el factor determinante del éxito en la industria de los préstamos financieros.

Durante los últimos años, muchos estudios se han realizado para superar el inconveniente principal de la computación blanda, que es la falta de *interpretabilidad*. Se muestra los diferentes enfoques utilizados para superar este problema. Entre los métodos utilizados están las reglas de extracción de las redes neuronales, el uso de redes híbridas. Métodos como la computación neuronal borrosa o la genética difusa o el uso de las redes neuronales no supervisadas como métodos de aprendizaje y auto-organización del mapeo de datos.

Se muestra los beneficios de la utilización de métodos híbridos para superar algunas limitaciones de los métodos individuales. Mediante el uso de sistema difusos y de inteligencia artificial junto con métodos como las redes neuronales, algoritmos genéticos y máquinas de soporte vectoriales se incrementa la transparencia del conjunto.

Otro tema es el clásico balance entre la precisión y la transparencia o la complejidad de los sistemas. Por ejemplo las redes neuronales son más precisas y menos transparentes que los modelos neuronales difusos y también los algoritmos genéticos son más precisos y menos transparentes que los sistemas difusos genéticos. Algunos otros métodos híbridos como el evolutivo neuronal borroso y los algoritmos de múltiples objetivos genéticos no han sido investigados todavía en estos modelos de cómputo híbridos.

Genetic Programming.

(Hussein Abdou, 2009)

La calificación de crédito ha sido ampliamente investigada en el área de finanzas, en general, y el sector bancario en particular. Recientemente, la programación genética (GP) ha atraído la atención tanto en el ámbito académico como en el empírico, hacia los problemas de crédito (figura 3.1). El objetivo principal es investigar la capacidad de la GP en este campo del control de créditos basados en la clasificación. Se propone como una extensión de los algoritmos genéticos y se inspira en la teoría de la evolución darwiniana para el análisis de modelos de control de crédito en los bancos egipcios del sector público.

El objetivo secundario es comparar la GP con el Análisis Probit (PA), la regresión logística y el peso de la evidencia medida (WOE), esta última, al parecer, una técnica ya olvidada en la investigación publicada. Dos criterios de evaluación se utilizan en este trabajo, para el promedio de clasificación correcta (CAC) que son: el criterio del tipo de coste estimado de errores de clasificación (CEM) con el criterio de coste de clasificación errónea diferente (MC), se dan con el fin de evaluar las capacidades de los modelos de calificación de crédito.

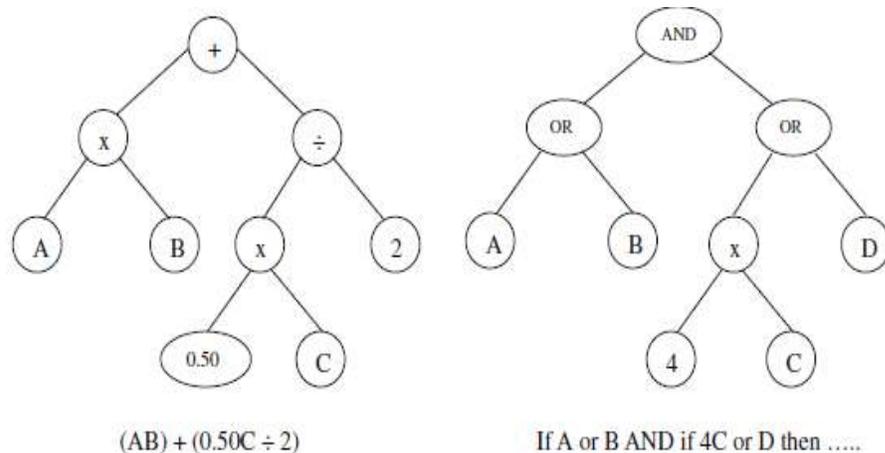


Figura 3.1. Dos ejemplos de GP usando operaciones matemáticas y condiciones (Hussein, 2009).

Los resultados hasta ahora revelan que la GP tiene la tasa más alta del CAC y el más bajo de CEM. Sin embargo, sorprendentemente, no existe

una norma clara para la medida de CAC en CEM con ratios más elevados de MC. Además, un análisis del conjunto de datos mediante mapas de Kohonen³ resulta prometedor para proporcionar más ideas visuales en grupos de racimo (*cluster groupings*).

Los resultados de la clasificación de todos los modelos ejemplos son mostrados en el cuadro 3.1. Resume las tasas ACC (exactitud, precisión) para las técnicas de calificación, es decir, WOE (Peso de la Evidencia), WOE1, PA (análisis Probabilístico), PA1, GPp (Programación Genética), and GPt. Se puede concluir a partir del cuadro que GPt tiene el mayor índice de ACC, que es 85,82%, entre todas las técnicas.

Todos los modelos predicen un mejor buen crédito y el mal crédito, a excepción de los modelos WOE. Los Modelos PA predicen un buen crédito mucho mejor que los modelos de WOE. Además, el más alto crédito clasificado correctamente malo fue de 98,30% para los modelos WOE, mientras que el mayor crédito bueno correctamente clasificado fue 91,89% para GPp. Como se muestra en el cuadro, el promedio del rendimiento de los modelos GP es mejor que el rendimiento de los modelos PA, pero mucho mejor que el rendimiento de las técnicas WOE.

Scoring model	Correctly classified results		
	G%	B%	ACC%
<i>Whole sample</i>			
WOE ^a	34.08	98.30	54.99
WOE ₁	33.25	98.30	54.44
PA	88.95	67.40	81.93
PA ₁	88.60	67.15	81.62
GP _p	91.89	65.45	83.28
GP _t ^b	91.07	74.94	85.82

^a Best model amongst all overall sample models based on EMC under MC ratio

^b Best model amongst all whole sample models based on ACC rate and EMC u

Cuadro 3.1. Resultados - Comparación Programación Genética
(Husein, 2009).

³ Método de Data Mining que proyecta una base de datos de N dimensiones a una de 2D y permite tener una idea de conjunto de toda la complejidad involucrada. Desarrollado por Kohonen en Holanda, 1989.

Neural nets versus conventional techniques.

(Hussein Abdou, y otros, 2008)

Las redes neuronales se han convertido en una de las herramientas más importantes para el manejo de la calificación del crédito. La calificación es considerada como un instrumento básico de tasación por parte de los bancos comerciales durante las últimas décadas.

El propósito es investigar la capacidad comparada de las redes neuronales, tales como las redes neuronales probabilísticas (figura 3.2), redes *feed-forward* de capas múltiples (figura 3.3), y las técnicas convencionales tales como análisis discriminante, análisis de PROBIT y regresión logística; en la evaluación de riesgo de crédito en los bancos egipcios que han optado por la aplicación de modelos de Calificación de crédito.

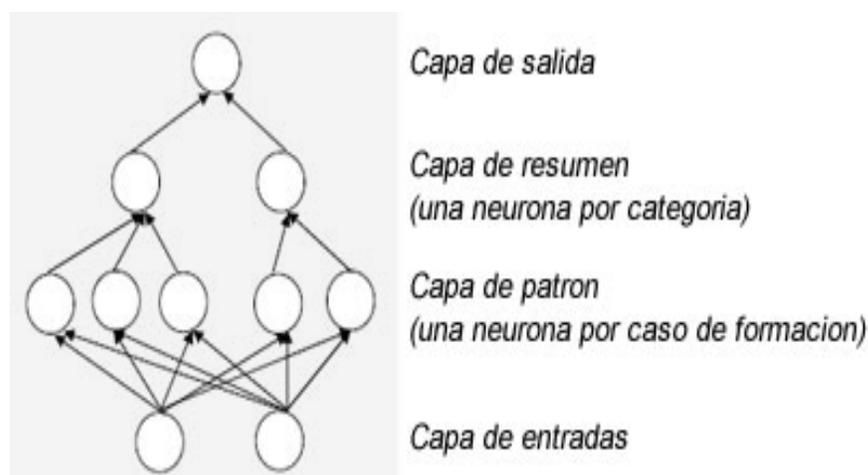


Figura 3.2. Estructura de PNN

(Hussein Abdou, y otros, 2008).

La tarea de campo para calificar el crédito, se realiza en un banco egipcio de préstamos personales que proporcionan los conjuntos de datos. Los resultados, hasta la fecha, revelan que las redes neuronales dieron un mejor promedio de clasificación correcta que las otras técnicas. Un análisis unidireccional de varianza y otras pruebas también se aplicaron y mostraron que hay algunas diferencias significativas entre las medias de las tasas de clasificación correcta, pertenecientes a diferentes técnicas.

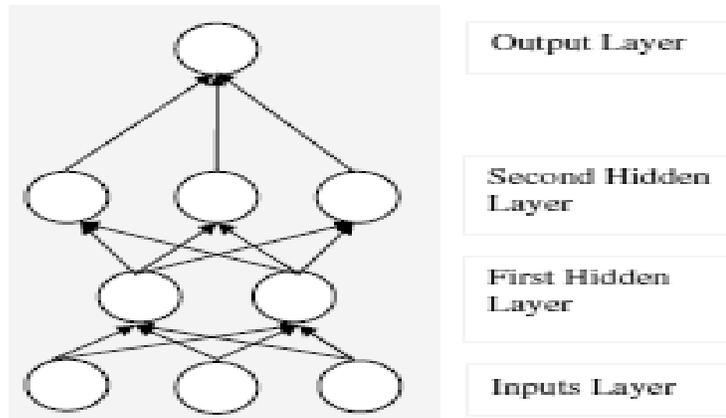


Figura 3.3. Estructura de MLF Nets
(Hussein Abdou, y otros, 2008)

El cuadro 3.2. resume los ratios promedios de la correcta clasificación de las técnicas convencionales DA (Análisis Discriminante), DA1, PA (Análisis Probabilístico), PA1, LR (Regresión Logística) and LR1, y los 5 mejores modelos de PNN (Red Neuronal Probabilística), los 5 mejores modelos de MLFN (Red Feed-forward Multi Capa) y los 5 mejores modelos de BNS (Mejor Red de búsqueda). Se puede concluir que el modelo LR tiene las mayores tasas medias de clasificación correcta, que es 87.95%, entre las técnicas convencionales. Mientras tanto el modelo PNN6 tiene la mayor tasa promedio de clasificación correcta, que es 96,21%, entre todas las técnicas.

Scoring model	Classification results (overall sample)		
	Good%	Bad%	Overall%
DA	85.91	89.19	86.75
DA ₁	85.91	89.86	86.92
PA	94.00	69.59	87.78
PA ₁	93.07	70.27	87.26
LR*	94.00	71.62	88.30
LR ₁	93.76	70.95	87.95
PNN ₁	97.23	87.84	94.84
PNN ₆ ****	98.61	89.19	96.21
PNN ₁₀	98.38	86.49	95.35
PNN ₁₃	98.61	85.14	95.18
PNN ₁₄	96.77	90.54	95.18
MLFN ₁	94.23	89.86	93.12
MLFN ₃ **	98.15	79.73	93.46
MLFN ₈ **	97.23	84.46	93.98
MLFN ₁₁	95.38	87.16	93.29
MLFN ₁₉	96.07	84.46	93.12
BNS-PNN	96.77	89.19	94.84
BNS ₇ -PNN***	97.46	87.84	95.01
BNS ₁₅ -MLFN-5N	96.07	91.22	94.84
BNS ₁₇ -MLFN-5N	97.46	86.49	94.66
BNS ₁₈ -MLFN-5N	95.61	92.57	94.84

- * Best conventional technique.
- ** Best MLFN.
- *** Best BNS with PNN.
- **** Best PNN and best of all techniques.

Cuadro 3.2. Resultados - Comparación Red Neuronal con técnicas convencionales (Hussein Abdou, y otros, 2008)

Todos los modelos predicen el “mejor buen crédito” que el “mal crédito”, a excepción de sólo dos modelos DA y DA1. Además, el mayor predictor malo fue 92,57% para BNS18-MLFN-5N, mientras que el más alto predictor bueno fue 98,61% para PNN6 y PNN13. La motivación de este trabajo fue evaluar el desempeño de determinadas redes neuronales, tales como PNNS y MLFNs (las mejores), frente a las técnicas convencionales, tales como DA, PA y LR, aplicadas en el sector bancario de Egipto.

Neural Nets and Genetic Programming versus Conventional Techniques

(Hussein Ali Hussein Abdou, 2009)

El *Credit Scoring* ha sido considerado como un instrumento de principal de evaluación de los bancos durante los últimos años, y ha sido ampliamente investigado en el área de finanzas en general y sectores de la banca en particular. En esta tesis, las principales metas y objetivos fueron: identificar las técnicas utilizadas actualmente en el proceso de evaluación de crédito bancario de Egipto, y para construir modelos de *Credit Scoring* para evaluar los préstamos bancarios personalizados. Además, de evaluar el impacto de la selección de la proporción de la muestra en la predicción de la capacidad de las técnicas de calificación avanzadas y las técnicas convencionales, con casos de estudio tanto para los bancos públicos y la banca privada, y para determinar las características claves que afectan la calidad de los préstamos personales (riesgo de incumplimiento).

Las etapas de la investigación estuvo compuesto, en primer lugar por una fase de investigación, incluyendo un primer piloto de estudio, entrevistas estructuradas y un cuestionario, y en segundo lugar, una fase de evaluación, incluyendo un análisis de dos diferentes conjuntos de datos de la banca privada y pública egipcia.

Las técnicas avanzadas, es decir, las redes neuronales (redes neuronales probabilísticas y redes feed-forward multicapas) y genéticos

programación, y las técnicas convencionales (medida del peso de la evidencia, análisis discriminante múltiple, el análisis probit y la regresión logística) se utilizaron para evaluar el riesgo de incumplimiento de crédito en los bancos egipcios. Además, de efectuar un análisis de los conjuntos de datos utilizando mapas de Kohonen que se llevó a cabo para proporcionar un entendimiento del conjunto de agrupaciones.

Desde la etapa de investigación, se encontró que los bancos públicos y en la inmensa mayoría los bancos privados en Egipto están utilizando enfoques de juicio en su evaluación de crédito. Desde la etapa de evaluación, las distinciones entre las técnicas convencionales y las técnicas avanzadas son claras. Se encontraron resultados que las técnicas avanzadas (como las redes neuronales y la programación genética) fueron superiores a las técnicas convencionales.

Dado que la tasa ACC se convirtió en un criterio importante en la evaluación de la clasificación de los modelos de scoring, esto fue importante para comparar los resultados de los diferentes modelos.

Scoring model	Overall sample		
	G %	B %	T %
WOE	51.27	100.00	63.68
WOE ₁	50.12	100.00	62.82
MDA	85.91	89.19	86.75
MDA ₁	85.91	89.86	86.92
PA	94.00	69.59	87.78
PA ₁	93.07	70.27	87.26
LR	94.00	71.62	88.30
LR ₁	93.76	70.95	87.95
PNN*	97.92	88.51	95.52
MLFN	96.07	85.81	93.46
GP _D	93.53	80.41	90.19
GP _T	94.46	82.43	91.39

*Best model amongst all models.

Cuadro 3.3. Resultados - Comparación Redes Neuronales, Programación Genética con técnicas convencionales - Bancos Privados (Hussein Abdou, 2009)

Como vemos en el cuadro 3.3. la PNN (Red Neuronal Probabilística) tiene el 95.52 de ACC, en el caso de estudio para los bancos privados de Egipto.

Scoring model	Overall sample		
	G %	B %	T %
WOE	34.08	98.30	54.99
WOE ₁	33.25	98.30	54.44
MDA	77.56	81.75	78.92
MDA ₁	78.03	81.51	79.16
PA	88.95	67.40	81.93
PA ₁	88.60	67.15	81.62
LR	89.31	67.15	82.09
LR ₁	89.19	67.64	82.17
PNN*	95.18	76.89	89.22
MLFN	89.89	75.67	85.26
GP _p	91.89	65.45	83.28
GP _t	91.07	74.94	85.82

*Best model amongst all whole sample models.

Cuadro 3.4. Resultados - Comparación Redes Neuronales, Programación Genética con técnicas convencionales – Bancos Públicos (Hussein Abdou, 2009)

Por último, en el cuadro 3.4. podemos ver que la PNN (Red Neuronal Probabilística) tiene el 89.22 de ACC, en el caso de estudio para los bancos públicos de Egipto.

Vertical bagging decision trees model.

(Defu Zhang, y otros, 2010)

En los últimos años, más y más gente en China, especialmente los jóvenes, comienzan a utilizar la tarjeta de crédito debido al cambio del concepto de consumo en ese país, por lo que el negocio de las tarjetas de crédito está creciendo rápidamente. Por lo tanto, es significativo el uso de herramientas eficaces, tales como los modelos de calificación de crédito, que sean creados para ayudar en la toma de decisiones a las empresas que participan en la emisión de las tarjetas de crédito.

En este trabajo se desarrolla un novedoso modelo de calificación crediticia, denominado modelo de empaquetamiento vertical, fundado en el manejo de árboles de decisiones (abreviado a VBDTM). El modelo es un método nuevo, diferente del tradicional empaquetamiento. El modelo VBDTM obtiene una agregación de los clasificadores mediante la combinación de atributos predictivos.

En el modelo VBDTM (figura 3.4), todas las muestras de las series de datos de aprendizaje y todos los atributos participan en el aprendizaje de

cada clasificador. En cambio con el método tradicional de empaquetado los clasificadores son entrenados solo con subconjuntos de muestras y cada clasificador tiene los mismos atributos. El VBDTM fue probado por dos bases de datos de crédito de la *UCI Machine Learning Repository*, y los resultados de los análisis muestran que el rendimiento del método propuesto es excepcional en la precisión de predicción.

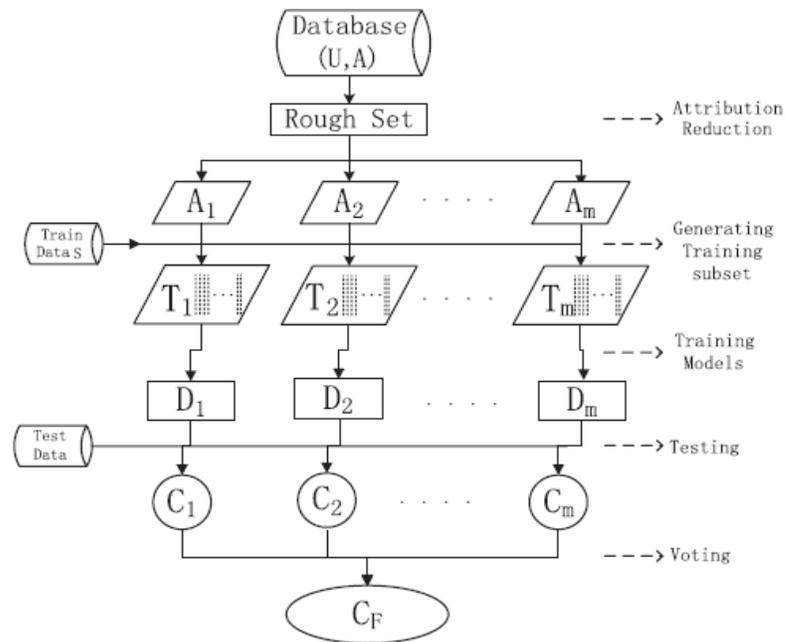


Figura 3.4. La estructura de VBDTM (Defu Zhang, 2010).

Comparison RSC and VBDTM on accuracy for two credit databases.

No.	German credit (%)		Australian credit (%)	
	RSC	VBDTM	RSC	VBDTM
1	79.5	80.8	87.68	90.73
2	79.7	83.5	88.12	92.03
3	80.4	81.3	87.54	92.90
4	80.0	81.5	88.26	91.89
5	79.2	81.6	88.26	92.03
6	79.7	82.2	88.12	91.89
7	79.0	81.4	87.54	92.46
8	80.4	81.6	88.12	91.45
9	79.0	80.6	87.54	92.75
10	80.4	81.9	87.83	91.60
Average (%)	79.73	81.64	87.90	91.97

Cuadro 3.5. Resultados - Comparación VBDTM con RSC (Defu Zhang, y otros, 2010)

De acuerdo al cuadro 3.5., la predicción de VBDTM es más exacta que RSC (Rough Set) para el crédito de Alemania, con una precisión promedio de 81,64% a 79,73%; VBDTM logra un aumento de 1,91 % del modelo RSC. Además, VBDTM supera a RSC con una precisión promedio de 91,97% a 87,90% para el crédito en Australia, mejorando en 4.07 %.

Comparison of different type model on accuracy for two credit databases.

Models		German credit (%)	Australian credit (%)
Single	C4.5	73.50	85.31
	MLP	74.68	86.78
	SVM	77.92	86.90
Rule extraction	CPAR	73.4	86.2
	Re-RX	80.54	
Two stages	RSC	79.73	87.90
	2SGP	79.49	90.17
Hybrid	HARA		89.28
	HCSM	79.68	89.45
Aggregation	Ensemble NN	75.82	87.71
	VBDTM	81.64	91.97

Cuadro 3.6. Resultados - Comparación VBDTM con otros modelos (Defu Zhang, y otros, 2010)

Los resultados del cuadro 3.6., confirman que VBDTM es más preciso y más robusto que otros tipos de modelo, como mejor modelo incluyendo C4.5, NN y SVM, modelos de dos estados incluyendo 2SGP y RSC y el método híbrido de HATC y HCSM, etc.

Support Vector machines and neural networks.

(Zan Huang, y otros, 2003)

El Análisis de calificación crediticia de las empresas ha atraído y atrae un enorme interés por parte de los investigadores y ello se ve reflejado en la creciente literatura publicada sobre el tema a nivel global. Algunos estudios recientes han demostrado que los métodos de la inteligencia artificial (IA) logran con frecuencia un mejor rendimiento que los métodos estadísticos tradicionales. Por ejemplo, recientemente, la inteligencia artificial (IA), en particular los basados en reglas de sistemas expertos, sistemas de razonamiento basados en casos; técnicas de aprendizaje de máquinas con redes neuronales, etc., se han utilizado para apoyar a dicho análisis.

Las técnicas de aprendizaje automático extraen el conocimiento de un conjunto de datos y “construyen” con él un modelo que proporciona diferentes representaciones explicativas de la estructura del conjunto de datos proporcionados para su aprendizaje. Se ha encontrado que la principal diferencia entre esta nueva aproximación al problema y la obtenible con los métodos tradicionales basados en probabilidad y estadística, residen en que estos métodos tradicionales le exigen al analista que imponga un modelo interpretativo de los datos (como Regresión Lineal, Análisis Factorial, etc.) antes de procesarlos por cálculos numéricos y estimar los parámetros y ajustar los datos de la observación.

Los nuevos modelos de cómputo, basados en IA, crean su propio modelo interpretativo por sí mismos y sin consultar con el analista. Este modelo interno es el núcleo del aprendizaje de la estructura del problema contenido de modo desmenuzado en el conjunto de datos de aprendizaje. Como resultado, mientras que en los métodos tradicionales las estructuras impuestas por el hombre, con los modelos estadísticos traídos de afuera, son relativamente simples y fáciles de interpretar, los nuevos modelos tienen un modelo interno de máquina muy oscuro y tan complicado que resulta muy difícil de explicarse para un ser humano. Afortunadamente no necesitamos comprenderlo en absoluto. ¡Hay que usarlo simplemente!

Es una técnica relativamente nueva para el aprendizaje de máquina con soporte vectorial (SVM), para el problema del aprendizaje eficaz y eficiente de manera que su modelo interno tenga mayor poder explicativo que el usualmente logrado con los métodos tradicionales.

Se utiliza retro-propagación de redes neuronales (BNN) como punto de referencia y la exactitud de predicción que se obtiene se halla alrededor del 80% para ambos métodos de BNN y SVM para los mercados de Estados Unidos y Taiwán. Sin embargo, se pudo apreciar sólo una ligera mejora respecto al SVM. Otra dirección de la investigación es el de

mejorar la interpretación de los modelos basados en IA. Han aplicado los resultados de las investigaciones recientes en la interpretación de modelo de red neuronal que modelan la importancia relativa de las variables de entrada financiera a los modelos de redes neuronales. Sobre la base de estos resultados, se llevó a cabo el análisis comparativo de mercados con las diferencias de los factores determinantes en los Estados Unidos y en Taiwán.

Prediction accuracies (LogR: logistic regression model, SVM: support vector machines, NN: neural networks)

	10-fold Cross-validation			Leave-one-out Cross-validation		
	LogR (%)	SVM (%)	NN (%)	LogR (%)	SVM (%)	NN (%)
TW I	72.97	79.73	75.68	75.68	79.73	74.32
TW II	70.27	77.03	75.68	70.27	75.68	74.32
US I	76.98	78.87	80.00	75.09	80.38	80.75
US II	75.47	80.00	79.25	75.47	80.00	75.68

Cuadro 3.7. Resultados – Comparación SVM con otros modelos
(Zan Huang, y otros, 2003)

El cuadro 3.7. resume la precisión de predicción de los tres modelos. Para fines de comparación. Se resume lo siguiente: el Soporte de Máquina Virtual logró los mejores resultados frente a la Red Neuronal (en algunos casos) y al Modelo de Regresión Logística.

En otro estudio realizado por Nnamdi y otros (2011) para la *World Academy of Science, Engineering and Technology 78*, indica que todos los modelos de RNA en este trabajo utilizan la propagación hacia atrás. Los resultados experimentales obtenidos muestran que la red neuronal supera al SVM (85,305% a 84%), mientras que el sistema de SVM requiere un tiempo de entrenamiento más corto que la RNA (0,08 a 452,03 seg.). Ver cuadro 3.8 y 3.9.

El tiempo de formación SVM sin embargo no incluye el rango de búsqueda de parámetros que es un tiempo de proceso largo y no se incluye aquí. Por lo tanto, se llegó a la conclusión que el RNA es mejor

debido a sus más altos índices de precisión y el tiempo mínimo de entrenamiento.

FINAL PARAMETERS OF THE CREDIT SCORING ARTIFICIAL NEURAL NETWORK LEARNER

	NN1	NN2
Number of Input Neurons	24	24
Number of Hidden Neurons	20	20
Number of Output Neurons	2	2
Weights Values Range	-0.35 to +0.35	-0.35 to +0.35
Learning co-efficient	0.00084	0.00052
Momentum Rate	0.44	0.909
Obtained Error	0.0079	0.00062
Performed iterations	4000	4000
Training time(s) [†]	452.03	452.78
Run Time(s) [†]	0.825x10 ⁻⁴	0.96x10 ⁻⁴
Training dataset accuracy rate	87.66%	92.09%
Testing dataset accuracy rate	73.60%	78.52%
Overall accuracy rate	80.63%	85.305%

Cuadro 3.8. Resultados - Aprendizaje de Red Neuronal (Nnamdi y otros, 2011)

FINAL PARAMETERS OF THE CREDIT SCORING SUPPORT VECTOR MACHINES LEARNER

Number of Features	24
Number of Classes	2
C Parameter Search Range	2 ⁻¹⁰⁰ – 2 ¹⁰⁰
γ Parameter Search Range	2 ⁻¹⁰⁰ - 2 ¹⁰⁰
C	32
γ	0.0156
V	5
Type of SVM used	C-SVM
Kernel	RBF
Training optimization time(s) [†]	0.08
Training dataset accuracy rate	84.4%
Testing dataset accuracy rate	83.6%
Overall accuracy rate	84%

Cuadro 3.9. Resultados - Aprendizaje de SVM (Nnamdi y otros, 2011)

Hybrid Classifier Using Neighborhood Rough Set and SVM (Ping Yao, 2009)

El desarrollo de modelos de Clasificación de crédito se ha ido convirtiendo en un tema de cómputo importante conforme la industria del crédito se fue tornando más y más competitiva. Por ello, la mayoría de

los modelos de Calificación de crédito han sido ampliamente estudiados en las áreas de probabilidad y estadística para mejorar la precisión de los modelos producidos durante los últimos años.

Se construye un modelo híbrido basado en soporte de máquina vectorial (SVM) para la calificación de crédito fundado en la evaluación de la puntuación del solicitante de crédito a partir de las características que puede mostrar como demandante. (1) Utilizando el criterio de vecindad peligrosa para seleccionar las características de entrada, (2) mediante la búsqueda en malla para optimizar los parámetros del núcleo RBF, (3) uso de las funciones híbridas óptimas de entrada y los parámetros del modelo para pasar a resolver el problema de Calificación de crédito aplicando validación cruzada por 10 veces, (4) comparar la exactitud del método propuesto con otros métodos.

Los resultados del experimento demuestran que el clasificador híbrido basado en la capacidad de crédito tiene una mejor puntuación en comparación con otros clasificadores híbridos. Con el fin de evaluar la eficacia del clasificador híbrido propuesto para el *Crédit Scoring*, los resultados también se comparan con aquellos que utilizan el análisis discriminante lineal, regresión logística y redes neuronales. Se puede concluir, en el cuadro siguiente, que el clasificador híbrido (conjunto de barrio peligroso y SVM) tiene una ligera mejor capacidad de crédito en términos de la tasa de exactitud y precisión.

TABLE VI. ACCURACIES WITH THE DIFFERENT METHODS FOR AUSTRALIAN AND GERMAN DATA SETS

method	Australian	German
linear discriminant analysis	0.8520	0.6600
logistic regression	0.8570	0.7240
neural networks	0.8683	0.7520
The proposed Hybrid classifier	0.8752	0.7660

Cuadro 3.10. Resultados - Comparación Clasificador Híbrido
(Ping Yao, 2009)

Modelo no lineal en RNA para clasificación.

(Martínez, 2000)

El principal objetivo de este trabajo es mostrar un tipo de redes neuronales denominadas redes neuronales basadas en unidades producto (RNUP) como un modelo no lineal que puede ser utilizado para la resolución de problemas de clasificación en aprendizaje. Proponen un método evolutivo en el que simultáneamente se diseña la estructura de la red y se calculan los correspondientes pesos. La metodología que presentamos se basa, por tanto, en la combinación del modelo no lineal RNUP y del algoritmo evolutivo; se aplica a la resolución de un problema de clasificación de índole económica, surgido del mundo de las finanzas.

Para evaluar el rendimiento de los modelos de clasificación obtenidos, comparan su propuesta con varias técnicas clásicas, como la regresión logística o el análisis discriminante, y con el clásico Modelo de perceptrón multicapa de redes neuronales basado en unidades sigmoides y el algoritmo de aprendizaje de retro-propagación (MLPBP). Los resultados del experimento demuestran que el RNUP tiene una mejor puntuación en comparación con las otras técnicas clásicas.

(Audante Néstor, 2007)

Esta tesis tiene por finalidad brindar a las empresas financieras (En especial a las PYMES Financieras) un sistema de Scoring, es decir, una forma de asignarle determinada calificación a los riesgos representados por clientes específicos, debido a que estas empresas presentan un alto índice de morosidad (tal y como lo demuestran los reportes de las cajas municipales de Ahorro y crédito y las cajas rurales) afectando tanto su rentabilidad como haciendo necesario la emisión de tasas de interés bastante altas para compensar el riesgo en el que incurren.

Las ventajas que supone el emplear un mecanismo de scoring son, entre otras: menor nivel de riesgo de la entidad, mejores niveles de recaudo de cartera y mejor utilización de los recursos, es decir, la aplicación de Scoring permite ante todo hacer más rentables los procesos.

3.3. Justificación del uso de la Red Neuronal Artificial

Como conclusión de la investigación del estado del arte, podemos afirmar que las **RNA (PNN y MLF)** están entre las dos (2) mejores técnicas avanzadas en la actualidad versus las técnicas convencionales. La comparación entre las técnicas revisadas, si es posible hacerlo a través de varias técnicas como: Matriz de confusión y II tipos de error, análisis incremental y análisis de comparación por resultados. En nuestro caso hemos revisado los resultados de Clasificación por precisión (Clasificación Accuracy) en los cuadros del 3.1. al 3.10., y se muestra en el siguiente cuadro:

Técnicas Convencionales	ACC %
WOE	54.99
PA	87.78
DA	86.92
LR	87.95
Técnicas Avanzadas	ACC %
PNN	96.21
GP	91.39
MLFN	93.98
VBDTM	91.97
SVM	86.90
HIBRIDO	89.65

Cuadro 3.11. Resultados ACC – Técnicas convencionales Vs. Técnicas Avanzadas (Fuente: Autor)

Las ventajas que se tiene de utilizar un RNA son:

- Resultados de las comparaciones de las técnicas avanzadas de *Credit Scoring*, que están en investigación.
- Herramientas libres y a bajo costo para construir RNAs como: MATLAB, SPSS, WEKA, EasyNN Plus, etc.
- Conocimiento y uso de las RNA en otros campos como: biología, ingeniería, economía, estadística, empresarial, negocios, medio ambiente, finanzas, manufacturación, medicina, militares, transporte y comunicaciones, etc.

- En nuestro medio aún no se usan las nuevas tendencias, salvo algunas aplicaciones de la banca que ya están comenzando a utilizar las RNA para: fraudes de tarjetas de crédito, concesión de préstamos, calificación de obligaciones, cotización de monedas extranjeras, mercado de valores, ratios financieros, etc.

CAPÍTULO IV

METODOLOGÍA DE LA INVESTIGACIÓN

En toda investigación, tecnológica o científica, es absolutamente fundamental la prueba, hasta un nivel prefijado de confianza, para cumplir con las proposiciones presentadas en la tesis. A continuación se indican los pasos metodológicos a seguir, para implementar el simulador convencional y la Red Neuronal Artificial (RNA), con la comparación respectiva de ambos.

4.1. Selección de muestras

Se consideró como universo del problema al conjunto total de clientes **potenciales** del Servicio de Microcrédito del Banco de la Nación. No era necesario averiguar su número ni su ámbito geográfico específico porque la **Teoría Estadística del Muestreo** solo requiere un subconjunto (probabilístico y homogéneo), o sea una pequeña **muestra representativa** de este conjunto para obtener resultados significativos (Murray, 1961).

Es importante la necesidad de entrevistar a los clientes reales del Servicio de Microcrédito, pero cabe señalar que este servicio **NO** existe todavía en el banco, por lo cual los datos con que se trabajaron fueron de carácter simulado (previamente obtenida de un sistema crediticio) y adaptada según la experiencia de un experto en el servicio de microcrédito.

La teoría del muestreo nos permite escoger una pequeña **muestra representativa** de clientes típicos de este Servicio simulados y generarlos y registrarlos en tablas de base de datos con resultados enteramente similares a los que se hubiesen obtenido si se hubiera analizado a clientes reales. La teoría del muestreo **garantiza** este resultado (hasta un 95% de confianza) si se cumplen tres condiciones:

- Que el conjunto de clientes de la muestra sea representativa y homogénea respecto a los clientes reales esperados.

- Que el número de clientes de la muestra sea calculada con la fórmula proporcionada por la teoría de muestreo estadístico.
- Que los miembros de la muestra sean elegidos aleatoriamente.

Estas condiciones fueron largamente cumplidas en el trabajo por lo que se procedió a calcular dicha muestra representativa con la fórmula de esta teoría.

La teoría del muestreo no hace depender pues de la magnitud de la población el tamaño de la muestra, sino que siguen paso a paso el procedimiento siguiente:

- Se estimó la **proporción** de clientes del universo total que posee la característica que estábamos investigando: la mora de los clientes. Esta proporción será nuestra variable **P**.

En este caso asumimos inicialmente que solo el 50 % de estos clientes logra evitar la mora al cabo de un año de conseguir el crédito. Por lo tanto el valor de la variable en cuestión fue:

$$P = 0.50.$$

- Luego se estimó la **variación V** que puede aceptarse en este porcentaje por razones de variabilidad inherente por causas varias no conocidas y lo situamos en $V = \pm 0.07$, es decir 7% pero con doble signo ya que la naturaleza no muestra preferencia por ninguna de las dos variaciones.
- El nivel de confianza asumido para el dato fue el estándar para este tipo de investigación, que es de 95 %, es decir que en 95 de cada 100 casos debe acertarse en la predicción y solo en el 5 % de los casos a lo sumo, fuera de los límites de variabilidad.
- Con este valor de confianza entramos a la **tabla Z** o normal para extraer el valor de la variable estadística $Z = 1.96$ que nos dice que la variabilidad que escogimos abarcaba 1.96 unidades de distancia respecto al valor central exacto. Entonces se calcula la desviación estándar S de la variable a investigar así:

$$S = \frac{V}{Z} = \frac{0.07}{1.96} = 0.0357$$

Luego se calcula el número de clientes, N, a ser generados así:

$$N = \frac{P(1-P)}{S \times S} + 1 = \frac{0.50(1-0.50)}{0.0357 \times 0.0357} + 1 = \mathbf{198}$$

Debíamos pues analizar a solo 198 clientes en cada una de las 2 muestras que se usaron para las pruebas (una de prueba con cómputo neuronal y la otra testigo con cómputo convencional) para que las pruebas fueran coherentes y rigurosas. Redondeando al próximo orden de magnitud, se obtuvo finalmente doscientos (200) clientes a generar para cada muestra de las pruebas.

4.2. Plan de desarrollo del Simulador Convencional y de la RNA

El Simulador convencional fue implementado utilizando la metodología Rational Unified Process - RUP (Kruchten, 2000), creando algunos artefactos necesarios para la investigación durante las disciplinas de captura de requerimientos, análisis, diseño, implementación y pruebas.

Para el entrenamiento de la RNA, se procedió a cargar la data y generar la topología de red para la RNA. Para esto se decidió utilizar la metodología CRISP_DM (Chapman, y otros, 2000) aplicando la minería de datos (Data Mining).

4.3. Comparación de los artefactos

Para comparar las soluciones del simulador convencional y de la RNA, se siguieron los siguientes pasos:

- Se establecieron los parámetros necesarios para la creación del programa de carga de la base de datos de clientes de Microcrédito, con el propósito de servir de base histórica de información real y también para los fines del entrenamiento del Programa RNA.
- Se creó el Programa Convencional de pronóstico de pagos de Clientes. Este programa es de tipo **Montecarlo**, es decir de sorteo aleatorio con funciones de probabilidad para la selección de los valores de cada registro. El programa se codificó con el lenguaje Java.

- Se instaló en una PC el programa de Computación neuronal **Easy NN Plus** del Distribuidor Neural Planner para la configuración y entrenamiento, por medio de un PATRÓN de aprendizaje, que era el perfil de pagos de los clientes con sus factores objetivos.
- Se corrió ambos programas de pronósticos (el convencional y el RNA) y se analizaron los resultados.

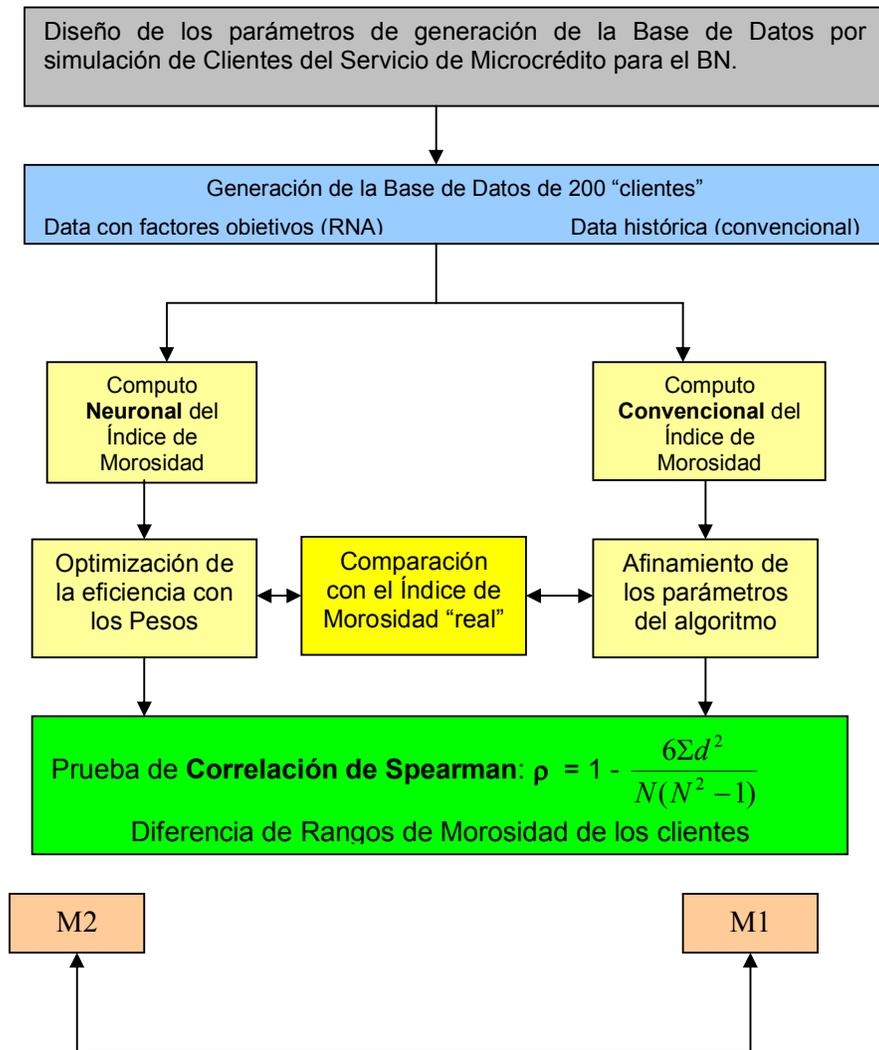
4.3.1. Método de Trabajo para la comparación

La obtención de la Evidencia Empírica (Morales, 2008) que avale los resultados en este trabajo, hasta un grado de confianza estadística de 95%, requiere de un proceso ordenado y jerárquico de pasos operativos que sucesivamente generan la base de datos de las muestras involucradas. Luego se efectúa en paralelo las corridas de computación neuronal y convencional respectivamente que generan los respectivos pronósticos de pagos de los clientes para los próximos seis (6) meses.

A continuación, se comparan sus respectivas eficiencias en el cálculo de los índices de morosidad predichas para los próximos seis meses de los clientes comparándolos con la base del historial acumulado de los seis meses anteriores. Esto se hace primero calculando el orden de rango de cada tipo de omisión de pagos (tabulándolos por tamaño) en la realidad de los clientes de los últimos seis meses y para cada uno de los dos métodos de predicción en prueba, para los próximos seis meses.

En seguida se calcula el Coeficiente de Correlación por Rangos de **Spearman** para estos tres historiales (real, predicho convencional, predicho RNA) tomándolos dos a dos.

Se completan los cálculos verificando la independencia mutua de estos tres tipos de registro de pagos de los clientes con la técnica de la Diferencia de Medias de **Pearson** y se sacan las conclusiones pertinentes. Ver esquema de la figura 4.1.



Prueba de Independencia por **Diferencia de Medias de Pearson**

Figura 4.1. Esquema de la Obtención de la Evidencia Empírica

4.3.2. Técnicas de recojo de datos y su tratamiento

Las técnicas de recojo de datos y su tratamiento se explican a continuación:

- Se procedió a aplicar el programa de cómputo neuronal de pronóstico de pagos y moras predicho para los próximos seis meses con sus factores objetivos, de los clientes de la base de datos. Paralelamente se hizo lo mismo aplicando el programa convencional, en Java, para que compute convencionalmente el mismo índice de moras de los clientes para los próximos seis meses. La salida de ambos

programas permitió conocer sus respectivos porcentajes de aciertos y fallas correspondientes.

- Se repitió el paso anterior una decena de veces con variantes, tales como ajustar o reconfigurar los “pesos” del programa neuronal o las probabilidades del programa convencional. Esto optimizó los logros de ambos programas y las listas de porcentajes obtenidas quedaron expeditas para el tratamiento estadístico.
- Una vez terminadas las pruebas de corrida de los programas Convencional y RNA se pudo tabular los datos en las siguientes ocho cuadros (4.1. al 4.8) de trabajo estadístico. A continuación se muestra un ejemplo:

OMISIONES	Nº CLIENTES	ORDEN DE RANGO
0	No= 40	3
1	N1= 55	2
2	N2 =70	1
3	N3 = 14	4
4	N4= 09	5
5	N5 = 08	6
6	N6 = 04	7

Cuadro 4.1. Datos Históricos **Reales** (6 meses anteriores)

OMISIONES	Nº CLIENTES	ORDEN DE RANGO
0	C0 = 23	4
1	C1= 64	1
2	C2 = 36	3
3	C3 = 53	2
4	C4 = 13	5
5	C5 = 05	7
6	C6 = 06	6

Cuadro 4.2. Pronóstico del programa **Convencional** (6 meses próximos)

OMISIONES	Nº CLIENTES	ORDEN DE RANGO
0	R0 = 41	3
1	R1 = 54	2
2	R2 = 67	1
3	R3 = 08	5
4	R4 = 19	4
5	R5 = 06	6
6	R6 = 05	7

Cuadro 4.3. Pronóstico del programa **RNA** (6 meses próximos)

Ahora ya se pueden hacer los cálculos estadísticos con el Coeficiente de Spearman:

- **Correlación de Spearman**, entre el orden de las Omisiones por tipo de programa de pronóstico empleado y el orden de las omisiones reales de los seis meses anteriores:

OMISIONES	Rango Convencional	Rango Real	D	D ²
0	4	3	1	1
1	1	2	-1	1
2	3	1	2	4
3	2	4	-2	4
4	5	5	0	0
5	7	6	1	1
6	6	7	-1	1
TOTALES				∑ d ² = 12

Cuadro 4.4. Correlación Convencional / Real

La fórmula de **Spearman** es: $P = 1 - \frac{6\sum d^2}{N(N^2 - 1)}$

Donde **P** es el coeficiente a calcular, **N** es el número de rangos de la muestra, y **d** es las diferencias entre rangos de ambas variables.

Reemplazando los valores obtenemos:

$$P = 1 - \frac{6 \times 12}{7 \times 48} = 0,78$$

Es decir existe una **fuerte** correlación directa por

rangos entre estas dos maneras de predecir el futuro de los pagos de estos clientes, tal como indica los argumentos teóricos de la

prueba de correlación para obtener la diferencia de rangos de la evidencia empírica, lo cual será validado en este trabajo.

Si se construye el cuadro de Rangos RNA / Real se tiene:

OMISIONES	Rango RNA	Rango Real	D	D ²
0	3	3	0	0
1	2	2	0	0
2	1	1	0	0
3	5	4	1	1
4	4	5	-1	1
5	6	6	0	0
6	7	7	0	0
TOTALES				$\Sigma d^2= 2$

Cuadro 4.5. Correlación RNA / Real

Reemplazando los valores en la fórmula de Spearman obtenemos:

$$P = 1 - \frac{6 \times 2}{7 \times 48} = 0,96$$

Es decir existe una **excelente** correlación directa

por rangos entre estas dos maneras de predecir el futuro de los pagos de estos clientes, tal como indica los argumentos teóricos de la prueba de correlación para obtener la evidencia empírica, lo cual será validado en este trabajo.

Construyendo el cuadro de Rangos RNA /Convencional se tiene:

OMISIONES	Rango RNA	Rango Convencio	D	D ²
0	3	4	-1	1
1	2	1	1	1
2	1	3	-2	4
3	5	2	3	9
4	4	5	-1	1
5	6	7	-1	1
6	7	6	1	1
TOTALES				$\Sigma d^2= 18$

Cuadro 4.6. Correlación RNA / Convencional

Reemplazando los valores en la fórmula de Spearman obtenemos:

$$P = 1 - \frac{6 \times 18}{7 \times 48} = 0,68$$
 Es decir existe una **débil** correlación directa por

rangos entre estas dos maneras de predecir el futuro de los pagos de estos clientes, tal como indica los argumentos teóricos de la correlación acerca de la **diferencia cualitativa** entre el RNA y la computación convencional con clara ventaja del primero.

- Acto seguido se estableció la **Prueba de Independencia**, también dos a dos, de estos tres conjuntos de datos:

Para ello se usó el Coeficiente de Diferencia de Medias de Pearson, que puede determinar si estos tres medios de conocer el comportamiento futuro de los clientes son independientes o no entre sí. Si son independientes entonces tiene sentido decir cuál de ellos es el mejor, si son equivalentes (no independientes) entonces puede usarse cualquiera de ellos sin ventaja alguna con ello. La expresión matemática de la **Prueba de Pearson** es:

$$Z = \frac{M1 - M2}{\sqrt{(S1^2 / n1) + (S2^2 / n2)}}$$

Donde:

M1 = Valor Medio del Índice de Morosidad con cómputo convencional

M2 = Valor Medio del Índice de Morosidad con cómputo neuronal

S1 = Desviación estándar correspondiente al caso M1

S2 = Desviación estándar correspondiente al caso M2

n1 = Número de rango de omisiones de la muestra para el cómputo convencional.

n2 = Número de rango de omisiones de la muestra para el cómputo neuronal.

Prueba de Independencia por **Diferencia de Medias de Pearson**:

Ahora se registran los datos de las diferencias. Se **construye** primero

dos Tablas de datos para Trabajo Estadístico a partir de las Tablas (4.1. a 4.6.):

OMISIONES	Nº con REAL	Nº con Convencional	Diferencia	Parámetros M1 y S1
0	40	23	17	Valor Medio: 15,43 Desviación Estándar 15.33
1	55	64	9	
2	70	36	34	
3	14	53	39	
4	09	13	4	
5	08	05	3	
6	04	06	2	
TOTALES	200	200		

NOTA: Las diferencias son absolutas.

Cuadro 4.7. Diferencias por tipo y cantidad de omisiones Real /
/Convencional

OMISIONES	Nº con REAL	Nº con RNA	Diferencia	Parámetros M2 y S2
0	40	41	1	Valor Medio: 3.43 Desviación Estándar 3.41
1	55	54	1	
2	70	67	3	
3	14	08	6	
4	09	19	10	
5	08	06	2	
6	04	05	1	
TOTALES	200	200		

NOTA: Las diferencias son absolutas.

Cuadro 4.8. Diferencias por tipo y cantidad de omisiones Real /
RNA

Con estos datos (Valor medio y Desviación Estándar) ya puede calcularse la fórmula de Pearson de **Diferencia de Medias de Pearson:**

$$Z = \frac{M1 - M2}{\sqrt{(S1^2 / n1) + (S2^2 / n2)}}$$

Este cálculo permite averiguar si ambos métodos (Convencional y RNA) son realmente **diferentes** (Independientes) o son **equivalentes** en cuanto a su poder de predicción (No independientes), en cuyo caso

daría lo mismo usar uno u otro método sin ventaja real de cualquiera de ellos.

Reemplazando los valores en la fórmula se tendrá:

$$Z = \frac{15.43 - 3.43}{\sqrt{(235 / 7) + (11.63 / 7)}} = 2,021$$

El **Z calculado** será pues = 2,021.

Por otra parte, ingresando a la Tabla estadística de la Distribución Normal de Gauss con el valor de Confiabilidad de 95% establecido, se extrae el valor **Z teórico** = 1,960.

Evidentemente pues el **Z calculado** (2,021) es mayor que el **Z Teórico** (1,960) y ello quiere decir que ambos métodos son INDEPENDIENTES.

El criterio estadístico a seguir para aceptar o rechazar los resultados será el usual para muestras distribuidas normalmente:

- **Z calculado > Z de Tabla => Resultado satisfactorio**
- **Z calculado <= Z de Tabla => Resultado NO satisfactorio.**

Está claro que el **Coefficiente de Spearman** permitió saber cuál de las dos formas de cómputo obtiene una mejor aproximación al orden del Índice de morosidad real de los clientes de la muestra y por lo tanto cual se aproximaba mejor a una predicción correcta de esta morosidad futura. Ese fue el RNA.

También está claro que el **Coefficiente de Pearson** permitió determinar que las dos formas de cómputo son realmente independientes entre sí, es decir las diferencias encontradas en sus resultados de cómputo se debían a diferencias cualitativas entre ambos métodos de cálculo y no al azar de las muestras tomadas. Esto garantizaba que los resultados hallados fuesen realmente objetivos y no meras especulaciones basadas

en datos numéricos o en comparaciones arbitrarias.

La inferencia final de ambas constataciones es **CONCLUYENTE**: “El Método RNA es mejor predictor del comportamiento crediticio de los clientes de Microcrédito que el Método Convencional y, al mismo tiempo ambos métodos son cualitativamente independientes. Esto será probado en el Capítulo VI de pruebas y experimentos numéricos”.

4.4. Determinación de requerimientos

4.4.1. Hardware

En este trabajo no se utilizó hardware neuronal debido a que excede la capacidad económica disponible para la investigación. Afortunadamente esto no era necesario realmente debido a que estos elementos de cómputo pueden simularse al 100% con Software por el simple expediente de representar cada “neurona” por un conjunto de funciones del tipo ya explicado en el apartado del Marco Teórico.

Se utilizó una (1) Laptop, para la simulación neuronal y la simulación convencional, ya que las tablas de Base de Datos no pasarán de 200 registros efectivos por lo que los requisitos de memoria y disco son mínimos. A pesar de ello se tuvo especial cuidado en asegurarse de la capacidad (tipo de procesador, memoria asignada, tipo de disco y espaciamiento), lo que eliminaba perturbaciones exógenas sobre las comparaciones que habrán de realizarse.

4.4.2. Software

Se eligió entre varios tipos de software neuronal disponibles en el mercado con arreglo a ciertos criterios tales como: precio, marca, historial operativo, capacidad y prestaciones.

Algunas de las Marcas disponibles encontradas fueron:

1) De la Distribuidora ALYUDA

- Alyuda Neuro Signal XL
- Forcaster
- Neuro Fusión.

2) De Attrasoft Inc.

- Predictor
- Predictor Pro

3) De CPC –X

- CPC – X Software

Un software neuronal muy atractivo, fue el Easy NN-Plus de “Distribuidor Neural Planner”; muy apropiado para nuestra investigación y por lo tanto se adoptó para este trabajo.

En cuanto al software convencional se usó el lenguaje Java EE. Este lenguaje de programación se usó para programar la simulación y generación de la base de datos de los clientes y para realizar el cómputo convencional del Índice de morosidad futuro de los clientes.

4.4.3. Sistemas Operativos

No hubo especial requerimiento del sistema operativo, ya que los programas involucrados incluyendo el software neuronal necesario para esta prueba, corrían perfectamente en Windows XP Release 3.0, que está disponible en el Banco de la Nación.

4.4.4. Base de Datos

La Base de datos utilizada fue Oracle 10g de 32 bits, constó de dos tablas idénticas de doscientos (200) registros cada una (indicado por la muestra calculada) y su descripción se halla pormenorizada en el Anexo B7. En la práctica se hizo extender hasta 1,000 registros durante las pruebas repetidas de doscientos (200) clientes cada una.

CAPÍTULO V

INGENIERÍA DE LOS ARTEFACTOS

5.1. Ingeniería del simulador convencional con RUP

Para la implementación de la solución del simulador convencional (primer aporte de la investigación) se desarrollaron las disciplinas de RUP que a continuación se detallan. Los Anexos que van desde B1 hasta B9 contienen los artefactos realizados para la simulación convencional. Se modeló con el Lenguaje de Modelamiento Unificado (UML) Versión 2.0 utilizando RSA (Rational Software Architect).

5.1.1. Captura de Requerimientos

Se efectuó un levantamiento de información a través de entrevistas (experto y gerente ejecutivo) y el análisis respectivo, para ver el interés, la factibilidad y la conveniencia de una futura implementación del servicio de microcrédito en el banco:

- Entrevista a un **Experto** en el Servicio de Microcrédito (Anexo A1)
- Entrevista al **Gerente Ejecutivo** sobre el nuevo Servicio y su importancia crítica para el banco (Anexo 2)
- Análisis **Documental** de la conveniencia de este Servicio (Anexo A3)
- Análisis de la **Magnitud y Restricciones** de este Servicio (Anexo A4).

En esta disciplina se analizó el Sistema de Prueba para ver sus actores (el investigador) y los casos de uso a implementarse. Se generó el prototipo, incluyendo la estructura de Menús (carga de datos, operaciones de cálculo, consultas y/o reportes).

Los artefactos creados en la captura de requerimientos fueron los siguientes:

- Actores de la prueba (Anexo B1)
- Diagrama de casos de uso de las pruebas (Anexo B2)
- Especificación de caso de uso “Simular Pago Convencional” (Anexo B3).

5.1.2. Análisis y Diseño Orientado a Objetos

Con los artefactos creados en la captura de requerimientos, se analizó y diseñó el caso de uso “Simular Pago Convencional”.

Además, de identificar los Nodos, Componentes, Clases y Pantallas para la construcción del programa convencional y las relaciones e interacciones mutuas entre estos componentes, bajo el enfoque orientado a objetos.

Los artefactos creados en el análisis y diseño fueron los siguientes:

- Diagrama de secuencia (Anexo B4)
- Diagrama de clases de diseño (Anexo B5)
- Algoritmo – Programa Convencional – Montecarlo (Anexo B6)
- Construcción de la base de datos y de las tablas (Anexo B7)
- Diagrama de Componentes del proceso (Anexo B8)
- Diagrama de Despliegue (Anexo B9).

5.1.3. Implementación

En primer lugar la herramienta tecnológica que se desarrolló en esta investigación, fue el algoritmo de Simulación Montecarlo (Morales, 2008) - Anexo C, fue aplicado para fines comparativos frente al RNA y solo a nivel de prototipo ya que no se está realizando un aplicativo formal para el Banco de la Nación (aplicación de usuario final).

Este prototipo realiza todas las operaciones básicas de la prueba comparativa planeada hasta el punto de mostrar fuera de toda duda si existen claras y significativas diferencias respecto a la capacidad predictiva del comportamiento de pagos de los futuros clientes de este nuevo servicio por el método tradicional algorítmico versus el método RNA aquí propuesto. El código fuente del programa convencional fue escrito en lenguaje Java EE. Los artefactos creados en esta disciplina son los siguientes: la codificación del software convencional y su instalación respectiva. Además de la implementación de la base de datos.

5.1.4. Pruebas

Las pruebas del simulador convencional, se basaron en la aplicación del método de Montecarlo con los doscientos (200) clientes seleccionados, según la especificación de caso de uso “Simular Pago Convencional”. Ver Anexo B3 y la figura 5.1.

Id.	Cod	Nombre	Pago Real	Pago Mx2	Pago Mx3	Pago Mx4	Pago Mx5	Suma Crecientes	Sm Crecientes	Diferencia	VA	
001	0006717546	CHOQUE AQUINO BARTOL PRIMO	0.00	0.00	0.00	163.73	0.00	0.00	6	0	2	2
002	0006716142	HUACA PAJARDO FELPE BARTHAÑO	0.00	0.00	85.51	85.51	0.00	0.00	4	0	4	4
003	0006727884	BORJA GUSPE VDA DE CASTAÑEDA FELICIA ZORADI	132.84	132.84	132.84	132.84	132.84	0	2	-2	2	
004	0006739688	MIAMI VAQUERO ARTURO	232.13	0.00	232.13	232.13	232.13	0.00	2	6	-8	3
005	0006731610	MURRY DE ORRIGO GLORIA NILDA	187.86	0.00	0.00	0.00	0.00	0.00	5	4	1	1
006	0006748542	LÓPEZ VÁSQUEZ CARLOS EDUARDO	0.00	238.19	0.00	238.19	238.19	0.00	3	4	-1	1
007	0006799248	OCAMPO SULLCAHUAMAN JESUS	132.43	132.43	132.43	132.43	132.43	0	0	0	0	0
008	0006798788	GUERRA DE VALENCIA ROSA FRANCISCA	184.74	184.74	184.74	184.74	184.74	0	6	-8	3	
009	0006809128	PAJELO CORDOVA ALICIA	289.81	0.00	289.81	289.81	289.81	0.00	2	4	-2	2
010	0006807388	BORJA CONDOR LEONIDO	0.00	82.37	0.00	0.00	0.00	0.00	5	1	4	4
011	0006854829	MIRABAL MUYA ALFREDO ALMO	83.13	83.13	83.13	83.13	83.13	0	0	0	0	0
012	0006863783	ORDÓÑEZ CORDOVA VDA DE CORDOVA LUDMILA	0.00	0.00	0.00	123.57	123.57	0.00	4	0	4	4
013	0006866023	OGAR BOPLOPOO CANDIDA ROSA	42.01	42.01	0.00	0.00	0.00	0.00	4	0	4	4
014	0006868323	AYENDANO DELA O DE SILVA JULIA	137.85	0.00	137.85	137.85	137.85	0.00	2	6	-8	3
015	0006882183	CANDIA ESCOBAR FRANCISCO	0.00	157.88	0.00	0.00	0.00	0.00	5	3	2	2
016	0041270888	MUÑOZ TRONCOSO RONALD	82.86	0.00	0.00	238.88	238.88	0.00	3	2	1	1
017	0028541437	CHAVEZ CERVANTES NILDA MERCEDES	0.00	388.80	0.00	388.80	388.80	0.00	3	0	3	3
018	0041284810	BOYZETA CRUZ DAVID BRAEL	0.00	281.28	0.00	281.28	281.28	0.00	3	6	-8	2
019	0006891181	MARTINEZ ALFARO AVELINO	187.83	187.83	187.83	187.83	187.83	0	3	-3	3	3
020	0028670316	CRUZ CARRASCO BLINDO	188.87	0.00	188.87	188.87	188.87	0.00	2	0	2	2
021	0028674877	SUCABACA ROQUE GREGORIO	0.00	285.75	285.75	0.00	0.00	0.00	4	2	2	2
022	0006898327	MENA PULACHE JUSTO ELIAS	111.80	111.80	0.00	0.00	0.00	0.00	4	1	3	3
023	0028683844	HUAMANI CAPAZA ANDRES NICOLAS	0.00	0.00	0.00	0.00	285.18	285.18	4	4	0	0
024	0028683887	QUIPETOONA JAIME JOHNNY	0.00	0.00	0.00	273.15	273.15	0.00	4	6	-1	1
025	0049327887	LOPEZ BARDALEZ ANGELY	282.80	282.80	0.00	0.00	0.00	0.00	4	2	2	2
026	0038824214	HUAMANI DE CONDORI DEMETRIA	337.81	337.81	337.81	337.81	337.81	0	0	0	0	0
027	0028479330	MADRID ZARATE JAIME	123.82	0.00	0.00	0.00	123.82	0.00	4	3	1	1
028	0002018487	YAGUINO PANDA MARCEL LEONIDAS	238.84	0.00	238.84	0.00	0.00	0.00	4	4	0	0
029	0040434121	ALVAREZ HERRERA JOSEPH JM	184.40	0.00	0.00	0.00	184.40	0.00	4	6	-1	1
030	0040881880	ALJAGA HINOJOSTROZA JOSE LUIS	287.86	287.86	0.00	0.00	0.00	0.00	4	6	-1	1
031	0002808238	LADINES MEJIA TITA ANTONIA	187.86	187.86	0.00	187.86	187.86	0.00	2	2	0	0
032	0028688913	VILLAR JAUREGUI ELNER ORLANDO	0.00	0.00	0.00	188.88	0.00	0.00	5	2	3	3
033	0018882885	MUNEZ MEJIA ESTANISLAO	0.00	0.00	0.00	0.00	0.00	171.15	5	4	1	1
034	0028687821	CANDIA CAUTI VDA DE CORDOVA OSDULA	0.00	138.23	0.00	0.00	0.00	0.00	5	6	0	0
035	0028488889	AUQUEJO DE SUAREZ DELA VIRGINIA	0.00	88.18	0.00	0.00	0.00	0.00	5	6	0	0
036	0008148718	EBICALANTE BENAVENTE FREDDY	180.25	0.00	180.25	0.00	0.00	0.00	4	6	-1	1

Figura 5.1. Data de Prueba – Simulador Convencional

5.2. Ingeniería de la Red Neuronal Artificial con CRISP-DM

En el trabajo de investigación (segundo aporte de la investigación) se ha utilizado la Metodología CRISP-DM, como guía para el desarrollo del componente Credit Scoring (Chapman, y otros, 2000). Nos ayudo a modelar nuestra RNA, durante las fases de: comprensión del negocio del microcrédito, la comprensión de los datos de entrada y salida, y su preparación, en el modelamiento de la topología de red, el entrenamiento de la red, su evaluación y el despliegue final.

El RNA es un producto comercial y por lo tanto no necesita construirse, siendo aplicado directamente en el proceso con el CRISP-DM. En la figura 5.2. se muestra el esquema del *Credit Scoring* con el RNA con la base de datos a interactuar.



Figura 5.2. Esquema del *Credit Scoring* con el RNA (Fuente: Autor)

5.2.1. Comprensión del Negocio

La calificación para un microcrédito requiere de la intervención de un analista de crédito en la recopilación y captación de información para el historial de crédito. La información es de tipo personal y cualitativo sobre el cliente, información que podría indicar factores de riesgo para la devolución del crédito concedido.

El procedimiento (ver la figura 2.2) para el otorgamiento del microcrédito (Salvador Rayo, y otros, 2010) pasa por las siguientes fases:

- **Promoción**

La fase comienza cuando se quiere captar clientes nuevos en una agencia o en el campo, para realizar una investigación de mercado y posteriormente la promoción del crédito.

El analista de crédito determina el procedimiento para informar y orientar al cliente sobre el microcrédito. Debido al desconocimiento por parte de los clientes al acceso al crédito, la entidad de micro finanzas realiza labores de asesoría financiera.

Si el cliente ha tenido créditos anteriores y tiene un historial, ya no pasa por esta fase. Las variables a tener en cuenta son: tiempo como cliente de la microfinanciera, créditos concedidos con anterioridad, créditos concedidos en el último año y créditos denegados con anterioridad. Para los clientes nuevos (sin créditos anteriores) su historial de pagos dependerá de la evaluación del analista de crédito, respecto a la ganancia mensual que tenga el cliente con su negocio, la cual debe ser verificada visitando al cliente y con referencias de otros entrevistados.

- **Evaluación**

En esta fase el analista de crédito determina un procedimiento para la evaluación de los expedientes de créditos a ser presentados al Evaluador de Créditos, que se encarga de aprobar o rechazar la solicitud. Las actividades que el analista de crédito realiza son tres:

- ✓ Análisis de la actividad económica que desarrolla el cliente. Las variables a tener en cuenta son: sector de actividad de la microempresa (comercio, agricultura, producción, servicios, etc.) y destino del microcrédito (capital de trabajo, activo fijo, etc.).
- ✓ Revisa y comprueba el comportamiento de pago del cliente en centrales de riesgo e historiales de crédito (si las tuviese).
- ✓ Inspección económico-financiera de la microempresa. En la mayoría de los créditos, el entorno macro económico que el cliente desarrolla la actividad principal es en el seno de la unidad familiar, donde se comprueba si posee un patrimonio.

El analista de crédito verifica la información patrimonial (ratios económicos y financieros), datos personales como: sexo, edad, estado civil y situación laboral y el tipo de garantía (declaración jurada, garantía real, aval, prenda, hipoteca, etc.).

- **Aprobación**

En esta fase el analista de crédito y el Evaluador de la entidad microfinanciera deciden aprobar o rechazar los expedientes de crédito presentados por el cliente.

Si la solicitud del préstamo termina por aprobarse, las variables a tener en cuenta serán: tipo de operación, moneda, número de cuotas del microcrédito, interés mensual y monto.

- **Desembolso**

El área de operaciones, revisa la carpeta de microcréditos aprobados, genera el cronograma de pagos, elabora el contrato para la firma del cliente. A su vez genera una cuenta para el abono respectivo. El cliente retirará el efectivo en la caja de la microfinanciera.

- **Recuperación del crédito**

El área de cobranzas de la microfinanciera, según el periodo pactado con el cliente, efectuará un seguimiento de pagos a través de: Información de control, visitas del analista de crédito, acciones de cobranzas, etc.

5.2.2. Compresión de los Datos

5.2.2.1. Clasificación de Variables para la RNA

Las variables para el *Credit Scoring* a implementar con la RNA, son los factores objetivos con los que se evalúa a un cliente (Hermosilla, 2005) y son los siguientes:

- **Veracidad** = $\text{N}^\circ \text{ de Documentos auténticos examinados} / \text{N}^\circ \text{ total de Documentos examinados}$.

- **Responsabilidad** = N° de Compromisos cumplidos / N° total de Compromisos asumidos.
- **Conducta ética** = N° de Clientes conformes / N° total de Clientes entrevistados por el banco.
- **Entorno social** = N° de Referencias favorables examinadas / N° total de Referencias examinadas.
- **Unidad familiar** = N° de familiares que participan (conyugue vale 2) / N° total de familiares directos.
- **Orden** = Nota del Evaluador después de visitar el negocio (Facilidad para ubicar documentos y general).
- **Conocimiento del negocio** = Nota del Evaluador con 5 preguntas (Precios, Mercados y Técnicas).
- **Ventajas competitivas** = Nota del Evaluador con 5 preguntas a clientes del negocio sobre satisfacción.
- **Ahorro** = Nota del Evaluador después de 5 preguntas sobre las cuentas de ahorro que tiene el negocio.
- **Autofinanciamiento** = Nota del Evaluador después de evaluar inversiones realizadas en el negocio.
- **Historial de pagos** = Suma de pagos de los 6 últimos meses / Suma máxima de pagos posibles.
- **Cifra de Mérito** = es un promedio ponderado de las notas obtenidas por el cliente al evaluar el banco sus factores personales de éxito. Para el caso de la comparación de la RNA con el Simulador Convencional, el valor obtenido se convertirá en un rango de omisiones (entre 0 y 6) que el cliente tendría (número de meses de mora), lo cual se distribuyó proporcionalmente entre los siete (7) rangos.

5.2.2.2. Determinación de Variables de Entrada y Salida.

A continuación se listan las variables de entrada y salida determinadas para la data a cargar en la RNA:

- **Variables de entrada**

Las variables de entrada determinadas son: veracidad, responsabilidad, conducta ética, entorno social, unidad, orden,

conocimiento del negocio, ventajas competitivas, ahorro, autofinanciamiento y historial de pagos.

- **Variables de salida**

La variable de salida determinada es: Cifra de mérito.

5.2.3. Preparación de los Datos

El objetivo de esta actividad es realizar un proceso de limpieza para descubrir comportamiento inadecuado en las variables de entrada y/o salida. En nuestro caso se afinaron las variables (factores objetivos) con el apoyo del experto de Microcrédito. A continuación se muestra la data preparada para el entrenamiento de la red. Ver figura 5.3.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	F
SecuenciaDni	Nombre	Veracidad	Responsabili	Conducta etic	Eritorno soci	Unidad famili	Orden	Conocimientc	Ventajas c	Ahorro	Auto financiero	Historial de r	Cifa de merito		
1	06717545	CHOGUE AQU	0.86	0.12	0.92	0.24	0.17	0.46	0.58	0.27	0.54	0.22	0.17	0	
2	06719142	HUACA FAJAR	0.02	0.85	0.32	0.04	0.2	0.5	0.01	0.18	0.72	0.54	0.33	0	
3	06727834	BORJAS QUISI	0.96	0.01	0.69	0.68	0.44	0.65	0.61	0.91	0.99	0.76	1	0	
4	06729559	MARIN VASQU	0.75	0.74	0.63	0.78	0.76	0.42	0.36	0.24	0.55	0.79	0.67	0	
5	06731610	ITURRY DE OR	0.84	0.39	0.96	0.13	0.75	0.32	0.42	0.87	0.24	0.32	0.17	0	
6	06748642	LOPEZ VASQU	0.99	0.55	0.99	0.53	0.41	0.61	1	0.57	0.35	0.43	0.6	0	
7	06799245	OCAMPO SULI	0.03	0.62	0.58	0.36	0.85	0.45	0.19	0.36	0.38	0.99	1	0	
8	06799788	GUERRA DE V	0.63	0.45	0.5	0.04	0.86	0.97	0.43	0.56	0.04	0.44	1	0	
9	06829129	PAJUELO COR	0.92	0.36	0.05	0.61	0.31	0.54	0.51	0.52	0.32	0.32	0.67	0	
10	06837399	BORJA CONDC	0.57	0.1	0.04	0.79	0.71	0.25	0.63	0.77	0.99	0.75	0.17	0	
11	06854829	MIRABAL MINA	0.53	0.08	0.08	0.95	0.48	0.92	0.47	0.59	0.18	0.71	1	0	
12	06863753	ORDO#EZ COF	0.73	0.28	0.47	0.19	0.66	0.8	0.74	0.36	0.59	0.08	0.33	0	
13	06868023	ODAR SOPLOF	0.52	0.74	0.3	0.39	0.93	0.54	0.59	0.45	0.74	0.09	0.33	0	
14	06866323	AVENDA#O DE	0.45	0.17	0.3	0.28	0.69	0.63	0.07	0.12	0.3	0.49	0.67	0	
15	06882193	CANDIA ESCO	0.37	0.61	0.23	0.87	0.18	0.14	0.35	0.04	0.72	0.25	0.17	0	
16	41270866	MU#OZ TRONC	0.33	0.88	0.87	0.12	0.05	0.06	0.62	0.11	0.88	0.83	0.6	0	
17	29541437	CHAVEZ CERV	0.74	0.17	0.22	0.46	0.02	0.25	0.48	0.96	0.16	0.25	0.5	0	
18	41284810	GOYZUETA CF	0.13	0.03	0.71	0.55	0.11	0.96	0.13	0.68	0.6	0.34	0.5	0	
19	06691191	MARTINEZ ALF	0.96	0.41	0.77	0.79	0.45	0.55	0.56	0.43	0.11	0.59	1	0	
20	29570315	CRUZ CARRAS	0.84	0.85	0.2	0.72	0.59	0.48	0.6	0.68	0.63	0.55	0.67	0	
21	29574977	SUCASACA R	0.85	0.35	0.5	0.59	0.82	0.53	0.5	0.59	0.3	0.63	0.33	0	
22	06698327	MENA PULACH	0.88	0.03	0.77	0.23	0.57	0.72	0.7	0.23	0.33	0.79	0.33	0	
23	29683644	MAMANI YILCA	0.65	0.95	0.35	0.48	0.03	0.45	0.36	0.21	0.83	0.71	0.33	0	
24	29683937	QUISPE TICON	0.3	0.75	0.53	0.24	0.84	0.53	0.33	0.05	0	0.5	0.33	0	
25	43327887	LOPEZ BARDA	0.3	0.24	0.19	0.98	0.49	0.47	0.21	0.05	0.52	0.67	0.33	0	
26	30824214	MAMANI DE CI	0.82	0.41	0.7	0.3	0.48	0.02	0.76	0.96	0.83	0.67	1	0	
27	25475032	MADRID ZARA	0.06	0.61	0.95	0.11	0.59	0.85	0.07	0.03	0.94	0.44	0.33	0	
28	02015467	YAGUNO PANI	0.5	0.36	0.38	0.81	0.86	0.73	0.28	0.66	0.88	0.98	0.33	0	
29	40434121	ALVAREZ HER	0.52	0.17	0.31	0.63	0.72	0.54	0.72	0.05	0.39	0.28	0.33	0	
30	40881950	ALIAGA HINOS	0.3	0.39	0.48	0.66	0.9	0.08	0.63	0.39	0.19	0.94	0.33	0	
31	02605236	LADINES MEJIL	0.16	0.22	0.47	0.36	0.49	0.92	0.02	0.15	0.46	0.8	0.67	0	
32	20668913	VILLAR JAURE	0.48	0.89	0.36	0.8	0.78	0.39	0.67	0.31	0.09	0.27	0.17	0	
33	16582635	NUÑEZ MEJIA	0.66	0.9	0.18	0.51	0.13	0.53	0.03	0.5	0.72	0.88	0.17	0	
34	25637821	CANDIA CAUTI	0.42	0.87	0.74	0.21	0.51	0.65	0.42	0.96	0.99	0.83	0.17	0	

Figura 5.3. Data de Prueba - Red Neuronal (Fuente: Autor)

5.2.4. Modelo de la Red

5.2.4.1. Topología de la Red

Se definió la topología de Red Perceptrón Multicapa. La red neuronal estará compuesta por tres capas, para nuestro caso con 21 neuronas:

- Entrada (11 nodos): Factores objetivos: Veracidad, responsabilidad, conducta ética, entorno social, unidad familiar, orden, conocimiento del negocio, ventaja competitiva, ahorro, auto financiamiento e historial de pagos.
- Oculta (7 nodos): sin nombres.
- Salida (1 nodo): cifra de mérito.

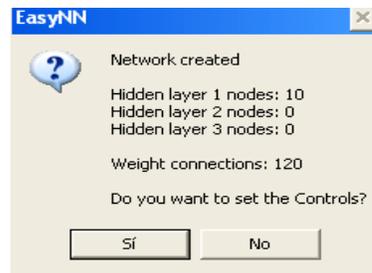


Figura 5.4. Creación de la Red Neuronal

Con el RNA “EasyNN–Plus” utilizamos un archivo Excel llamado Data_entrenamiento.xls, con los datos para el entrenamiento de la red. El archivo será importado al aplicativo como “Entrenamiento_RNA.tvq”.

A continuación en la figura 5.5. se muestra el diseño de la red neuronal creada con los datos de los clientes y sus factores de éxito (capa de entrada), además de la cifra de mérito (capa de salida) a obtener después de la calificación.

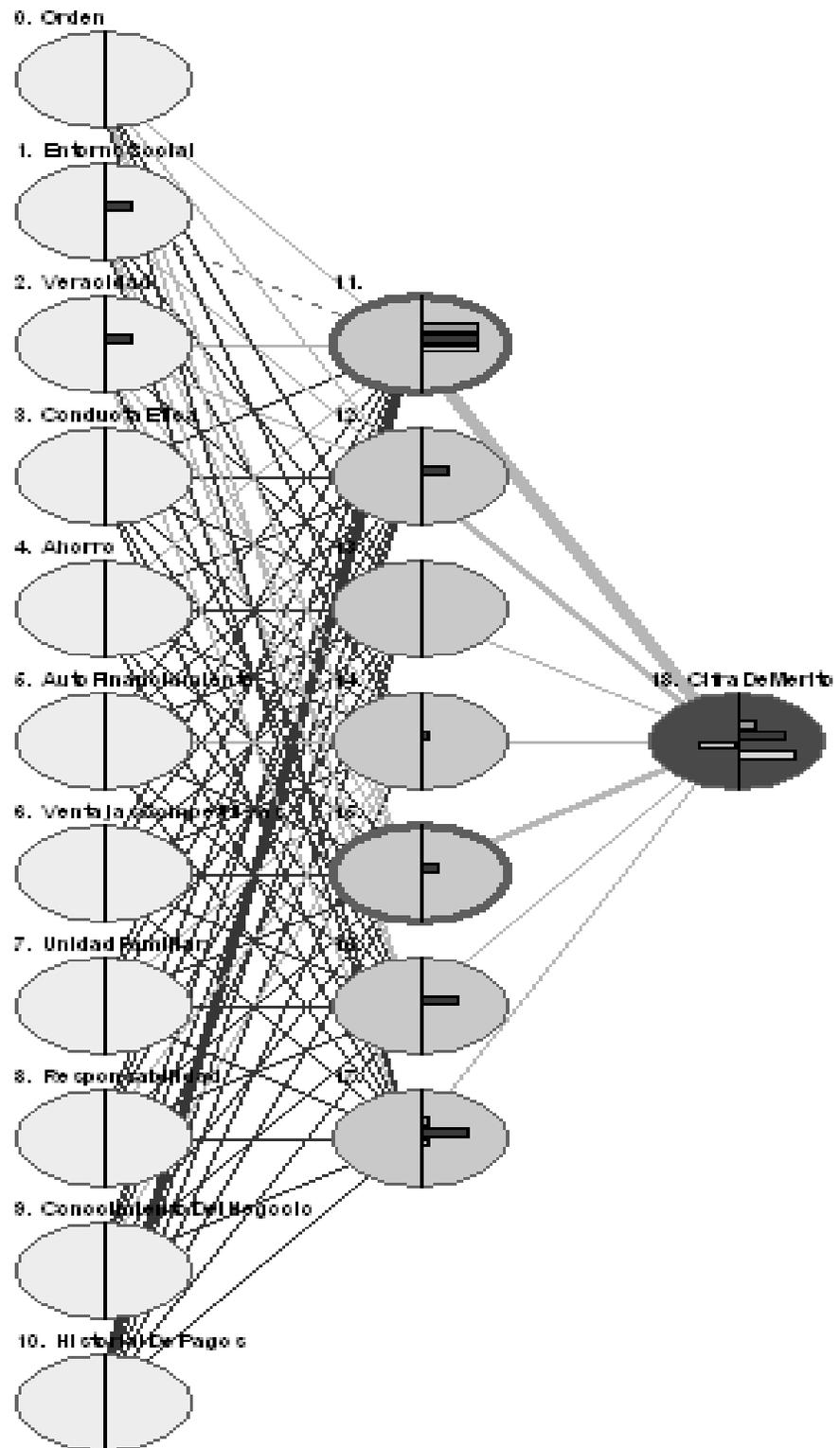


Figura 5.5. Topología de la Red Neuronal

5.2.4.2. Pruebas

Las actividades de pruebas se realizan para medir el rendimiento en función del aprendizaje. En las figuras 5.6. y 5.7. se muestran los controles configurados.

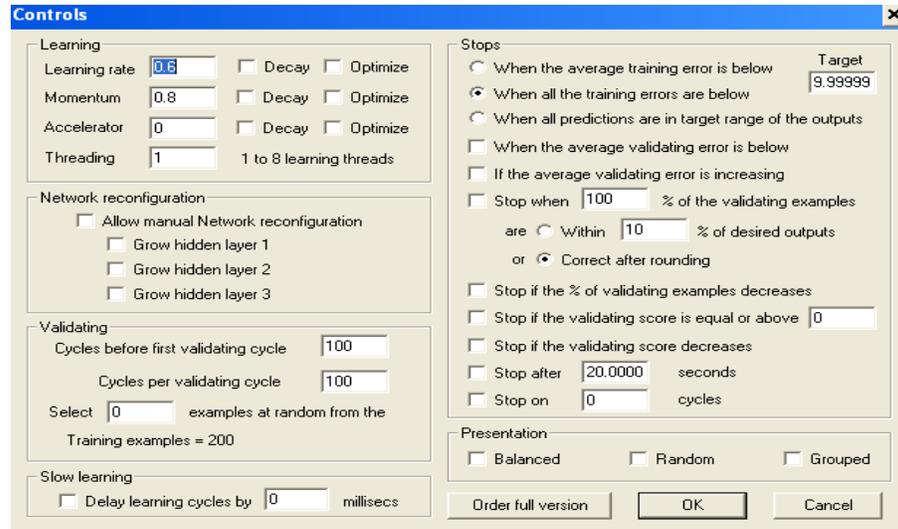


Figura 5.6. Configuración de Controles para el Entrenamiento

Es necesario configurar los controles para el aprendizaje de la red. Para nuestro caso la red se va a entrenar con la siguiente configuración:

- Margen de error al 0.00001.
- Ratio de aprendizaje: 0.6

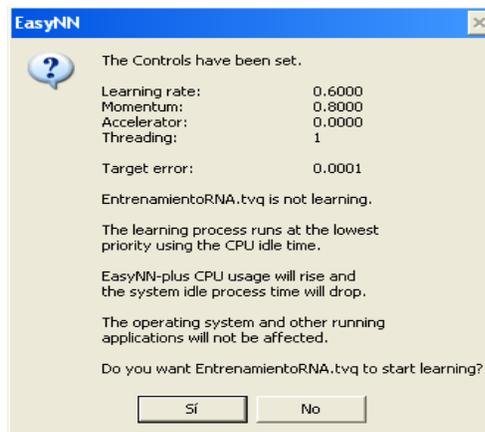


Figura 5.7. Controles Configurados

EasyNN-Plus es una aplicación de emulación de redes neuronales multicapas basada en algoritmo de aprendizaje de retropropagación y la función sigmoide para la activación de una neurona. En este caso la red será entrenada y calculará los pesos sinápticos de cada neurona

ciclando n veces hasta alcanzar el error medio requerido. Ver figura 5.8.

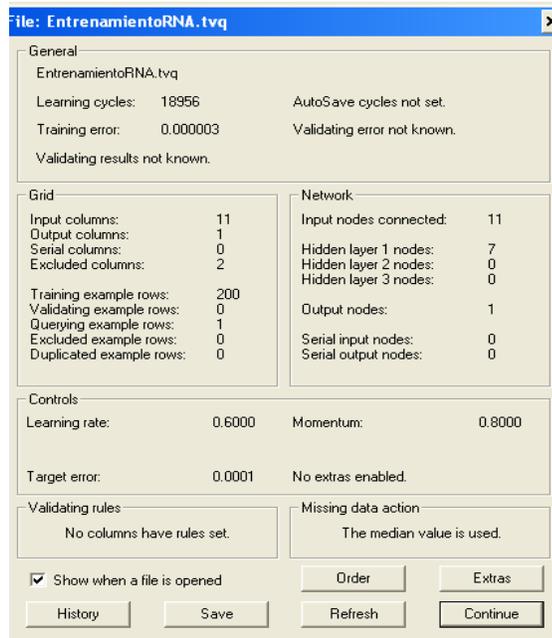


Figura 5.8. Resultado del Entrenamiento

En la figura 5.9. se muestra la curva de aprendizaje, de la red, luego de ciclar **18,956 veces**. La figura 5.10. muestra la data entrenada con el RNA.

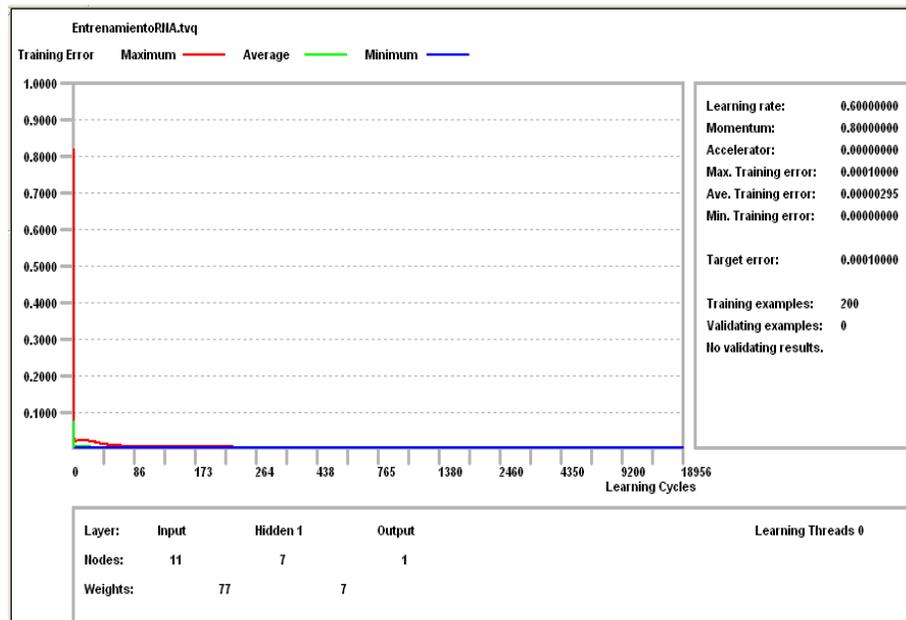


Figura 5.9. Curva de Aprendizaje de la Red

Dni	Nombre	Veracidad	Responsabi+	Conducta e+	Entorno _e+	Unidad _fa+	Orden	Conocimient+	Ventajas+	Ahorro	Auto_finant+	Historial +	Cifra de
001	067117545	CHOQUE AQUI	0.8600	0.9200	0.2400	0.1700	0.4600	0.5800	0.2700	0.5400	0.2200	0.1700	5
002	067119142	HUACA FAJAR	0.8500	0.3200	0.0400	0.2000	0.5000	0.0100	0.1800	0.7200	0.5400	0.3300	4
003	067278834	BORJAS QUIS	0.9600	0.6900	0.6800	0.4400	0.6500	0.6100	0.9100	0.9900	0.7600	1.0000	0
004	067295959	MARIN VASQU	0.7500	0.6300	0.7800	0.7600	0.4200	0.3800	0.2400	0.5500	0.7900	0.6700	2
005	067316110	ITURRY DE O	0.8400	0.9600	0.1300	0.7500	0.3200	0.4200	0.8700	0.2400	0.3200	0.1700	5
006	067485442	LOPEZ VASQU	0.9900	0.9900	0.5300	0.4100	0.6100	1.0000	0.5700	0.3500	0.4300	0.5000	3
007	067992445	MARIN VASQU	0.0300	0.5800	0.3800	0.8500	0.4500	0.1900	0.3600	0.3800	0.9900	1.0000	0
008	067997988	GUERRA DE V	0.6300	0.5000	0.0400	0.8600	0.9700	0.4300	0.5600	0.0400	0.4400	1.0000	0
009	068291249	PAJUELO COR	0.9200	0.0500	0.6100	0.3100	0.5400	0.5100	0.5200	0.3200	0.3200	0.6700	2
010	068373999	BORJA CONDO	0.5700	0.0400	0.7900	0.7100	0.2500	0.8300	0.7700	0.9900	0.7500	0.1700	5
011	068548299	MIRABAL HIN	0.5300	0.0800	0.9500	0.4800	0.9200	0.4700	0.5900	0.1800	0.7100	1.0000	0
012	068663753	ORDOÑEZ COR	0.7300	0.4700	0.1900	0.6600	0.8000	0.7400	0.3600	0.5900	0.0800	0.3300	4
013	068666023	ODAR SOPLOP	0.5200	0.3000	0.3900	0.9300	0.5400	0.5900	0.4500	0.7400	0.0900	0.3300	4
014	068666323	AYENDA#O DE	0.4500	0.3000	0.2800	0.6900	0.6300	0.0700	0.1200	0.3000	0.4900	0.6700	3
015	068882193	CAMDIA ESCO	0.3700	0.2300	0.8700	0.1800	0.1400	0.3500	0.0400	0.7200	0.2500	0.1700	5
016	412708666	MU#02 TROMC	0.3300	0.8700	0.1200	0.0500	0.0600	0.6200	0.1100	0.8800	0.8300	0.5000	3
017	295414377	CHAVEZ CERV	0.7400	0.2200	0.4600	0.0200	0.2500	0.4800	0.9600	0.1600	0.2500	0.5000	3
018	412848110	GOYZUETA CR	0.1300	0.7100	0.5500	0.1100	0.9600	0.1300	0.6800	0.6000	0.3400	0.5000	3
019	066911191	MARTINEZ AL	0.9600	0.7700	0.7900	0.4500	0.5500	0.5600	0.4300	0.1100	0.5900	1.0000	0
020	068666323	CRUZ CARRAS	0.8400	0.2000	0.7200	0.5900	0.4800	0.6000	0.6800	0.6300	0.5500	0.6700	2
021	295749777	SUCASACA RO	0.8500	0.5000	0.5900	0.8200	0.5300	0.5000	0.5900	0.3000	0.6300	0.3300	4
022	066983277	MEMA PULACH	0.8800	0.7700	0.2300	0.5700	0.7200	0.7000	0.2300	0.3300	0.7900	0.3300	4
023	295836444	MAMANI VILC	0.6500	0.3500	0.4800	0.0300	0.4500	0.3600	0.2100	0.8300	0.7100	0.3300	4
024	295883937	QUISPE TICO	0.3000	0.5300	0.2400	0.8400	0.5300	0.3300	0.0500	0.0000	0.5000	0.3300	4
025	067278834	LOPEZ BARDA	0.3000	0.1900	0.9800	0.4900	0.4700	0.2100	0.0500	0.5200	0.6700	0.3300	4
026	308242144	MAMANI DE C	0.8200	0.7000	0.3000	0.4800	0.0200	0.7600	0.9600	0.8300	0.6700	1.0000	0
027	067191422	MADRID ZARA	0.0600	0.9500	0.1100	0.5900	0.8500	0.0700	0.0300	0.9400	0.4400	0.3300	4
028	020154677	YAGUÑO PAND	0.5000	0.3600	0.8100	0.8600	0.7300	0.2800	0.6600	0.8800	0.9800	0.3300	4
029	404341211	BORJAS QUIS	0.5200	0.3100	0.6300	0.7200	0.5400	0.7200	0.0500	0.3900	0.2800	0.3300	4
030	067485442	ALIAGA HINO	0.3000	0.4800	0.6600	0.9000	0.0800	0.8300	0.3900	0.1900	0.9400	0.3300	4
031	026052388	LADINES MEJ	0.6000	0.4700	0.3600	0.4900	0.9200	0.0200	0.1500	0.4600	0.8000	0.6700	2
032	206689133	VILLAR JAUR	0.4800	0.3600	0.8000	0.7800	0.3900	0.6700	0.3100	0.0900	0.2700	0.1700	5
033	165826355	MUNEZ MEJIA	0.6600	0.1800	0.5100	0.1300	0.5300	0.0300	0.5000	0.7200	0.8800	0.1700	5
034	256378211	CAMDIA CAUT	0.4200	0.7400	0.2100	0.5100	0.6500	0.4200	0.9600	0.9900	0.8300	0.1700	5
035	068548299	AUSEJO DE S	0.1100	0.4900	0.0900	0.0700	0.0600	0.9800	0.0800	0.6700	0.2900	0.1700	5

Figura 5.10. Data de Entrenamiento – Con Cifra de Mérito (Fuente: Autor)

El grado de importancia según el experto en microcréditos para los factores objetivos (pesos asignados), para el entrenamiento realizado fueron los siguientes:

1. Historial de pagos	20
2. Conocimiento del negocio	20
3. Responsabilidad	10
4. Unidad familiar	10
5. Ventajas competitivas	10
6. Autofinanciamiento	10
7. Ahorro	5
8. Conducta ética	5
9. Veracidad	5
10. Entorno social	3
11. Orden	2
TOTAL	100

En la siguiente figura 5.11 se ve el grado de importancia de cada uno de los factores objetivos que se usan para evaluar a un cliente.

EntrenamientoRIA.tvq 18956 cycles. Target error 0.0001 Average training error 0.000003

The first 11 of 11 Inputs in descending order.



Figura 5.11. Grado de Importancia de los factores Objetivos

En la siguiente figura 5.12 se ve el grado de sensibilidad de cada uno de los factores objetivos que se usan para evaluar a un cliente.

EntrenamientoRNA.tvq 18956 cycles. Target error 0.0001 Average training error 0.000003

The first 11 of 11 Inputs in descending order. Output column 13 CifraDeMerito

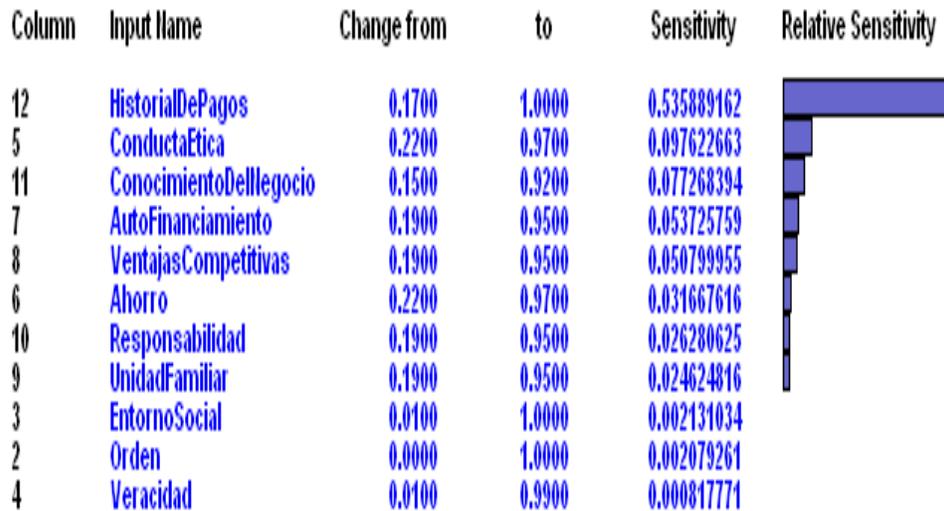


Figura 5.12. Grado de Sensibilidad de los Factores Objetivos

Por último se ve en la figura 5.13 el orden de importancia de cada uno de los factores objetivos de menor a mayor.

EntrenamientoRNA.tvq 18956 cycles. Target error 0.0001 Average training error 0.000003

The first 11 of 11 Inputs in ascending order.



Figura 5.13. Orden de Importancia de los Factores Objetivos

A continuación, la figura 5.14 muestra el resultado de la corrida del entrenamiento del RNA, con el número de omisiones de los clientes por grupos.

OMISIONES	No. CLIENTES
0	36
1	9
2	32
3	24
4	32
5	67
6	0

Figura 5.14. Resultado del Entrenamiento

5.2.5. Evaluación de la red neuronal

Una vez entrenada la red neuronal, se procede a verificar que lo aprendido por esta puede servir para predecir los resultados de las cifras de mérito a obtener de cada uno de los clientes. Para la evaluación de la red se crearon datos de prueba, con la misma información de los datos de entrenamiento, sin la cifra de mérito (ver figura 5.16), para validar que los resultados sean los mismos. Con esto quedaría demostrado que la red ha aprendido. Utilizaremos un archivo llamado Data_Prueba.xls.

La figura 5.15 muestra el resultado de la validación efectuada a la RNA la cual es igual al resultado de la figura 5.14.

OMISIONES	NO. CLIENTES
0	36
1	9
2	32
3	24
4	32
5	67
6	0

Figura 5.15. Resultado de Validación de la Prueba

5.2.6. Despliegue o Implementación

El componente neuronal fue implementado con los requerimientos mencionados en el punto 4.4. del capítulo IV (Determinación de requerimientos).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Secuencia/Dni	Nombre	Veracidad	Responsabil/Conducta etic	Entorno soci/Orden	Unidad form/Orden	Compromisment	Verdajaz c/Ahoro	Auto financiar	Historial de r	Cifra de merito			
2	001	06717545	CHODLE AGUI	0.85	0.12	0.92	0.24	0.17	0.45	0.58	0.27	0.54	0.22	0.17
3	002	06719142	HUACA FAJAR	0.02	0.85	0.32	0.04	0.2	0.5	0.01	0.18	0.72	0.54	0.33
4	003	06727834	BORJAS OLIVE	0.95	0.01	0.69	0.68	0.44	0.65	0.61	0.91	0.99	0.76	1
5	004	06729659	MARIN VASOUL	0.75	0.74	0.63	0.78	0.76	0.42	0.38	0.24	0.95	0.79	0.67
6	005	06731610	ITURRY DE OR	0.84	0.39	0.95	0.13	0.75	0.32	0.42	0.87	0.24	0.32	0.17
7	006	06746542	LOPEZ VASOUL	0.99	0.65	0.99	0.53	0.41	0.61	1	0.57	0.35	0.43	0.5
8	007	06795245	OCAÑO SULLI	0.03	0.62	0.98	0.38	0.85	0.45	0.19	0.36	0.38	0.99	1
9	008	06799798	GUERRA DE V	0.63	0.45	0.5	0.04	0.86	0.97	0.43	0.95	0.04	0.44	1
10	009	06829129	PAJUELO COR	0.36	0.36	0.05	0.61	0.31	0.54	0.51	0.52	0.32	0.32	0.67
11	010	06837399	BORJA CONDO	0.57	0.1	0.04	0.79	0.71	0.25	0.63	0.77	0.99	0.75	0.17
12	011	06854829	MIRABAL MINA	0.53	0.8	0.08	0.95	0.48	0.92	0.47	0.69	0.18	0.71	1
13	012	06863753	ORDÓÑEZ COF	0.73	0.26	0.47	0.19	0.66	0.8	0.74	0.36	0.59	0.08	0.33
14	013	06866023	ODAR SORLOI	0.52	0.74	0.3	0.39	0.93	0.54	0.59	0.45	0.74	0.09	0.33
15	014	06866323	AVENDANO DE	0.45	0.17	0.3	0.26	0.69	0.63	0.07	0.12	0.3	0.49	0.67
16	015	06882193	CANDIA ESCO	0.37	0.61	0.23	0.87	0.18	0.14	0.35	0.04	0.72	0.25	0.17
17	016	06882956	MUNOZ TRONC	0.33	0.88	0.87	0.12	0.05	0.06	0.62	0.11	0.88	0.83	0.5
18	017	29541437	CHAVEZ CERN	0.74	0.17	0.22	0.46	0.02	0.25	0.82	0.96	0.16	0.25	0.5
19	018	41284810	GOYUETA CF	0.13	0.03	0.71	0.55	0.11	0.96	0.13	0.68	0.6	0.34	0.5
20	019	06881191	MARTINEZ ALF	0.95	0.41	0.77	0.79	0.45	0.55	0.95	0.43	0.11	0.59	1
21	020	29570315	CRUZ CARRAS	0.84	0.85	0.2	0.72	0.59	0.48	0.6	0.68	0.63	0.55	0.67
22	021	29574977	SUCASACA RC	0.85	0.36	0.5	0.59	0.82	0.53	0.7	0.59	0.3	0.63	0.33
23	022	06886327	MENA PULACH	0.88	0.03	0.77	0.23	0.57	0.72	0.7	0.23	0.33	0.79	0.33
24	023	29583544	MAMAN YILCA	0.65	0.95	0.35	0.48	0.03	0.45	0.36	0.21	0.83	0.71	0.33
25	024	29583837	OLISPE TICON	0.3	0.75	0.53	0.24	0.84	0.53	0.33	0.05	0	0.5	0.33
26	025	43327857	LOPEZ BARDOA	0.3	0.24	0.19	0.98	0.49	0.47	0.21	0.05	0.52	0.67	0.33
27	026	38024214	MAMAN DE CI	0.82	0.41	0.7	0.3	0.48	0.02	0.76	0.96	0.83	0.67	1
28	027	25475032	MADRID ZARA	0.05	0.61	0.95	0.11	0.89	0.85	0.07	0.03	0.94	0.44	0.33
29	028	00015467	YAGUINO PAMI	0.5	0.36	0.38	0.81	0.86	0.73	0.28	0.66	0.85	0.96	0.33
30	029	40434121	ALVAREZ HER	0.52	0.17	0.31	0.63	0.72	0.54	0.72	0.05	0.39	0.28	0.33
31	030	40881950	ALJAGA HINOS	0.3	0.39	0.48	0.66	0.9	0.08	0.83	0.39	0.19	0.94	0.33
32	031	00805238	LADINES MEL	0.6	0.22	0.47	0.36	0.49	0.92	0.02	0.15	0.46	0.8	0.67
33	032	20666813	VILLAR JAURE	0.48	0.89	0.36	0.8	0.78	0.39	0.67	0.31	0.09	0.27	0.17
34	033	16582635	MUNEZ MEJIA	0.85	0.9	0.18	0.51	0.13	0.53	0.03	0.5	0.72	0.88	0.17
35	034	25637821	CANDIA CAUTI	0.42	0.87	0.74	0.21	0.51	0.65	0.42	0.96	0.99	0.83	0.17

Figura 5.16. Data de Prueba de validación – Sin Cifra de Mérito

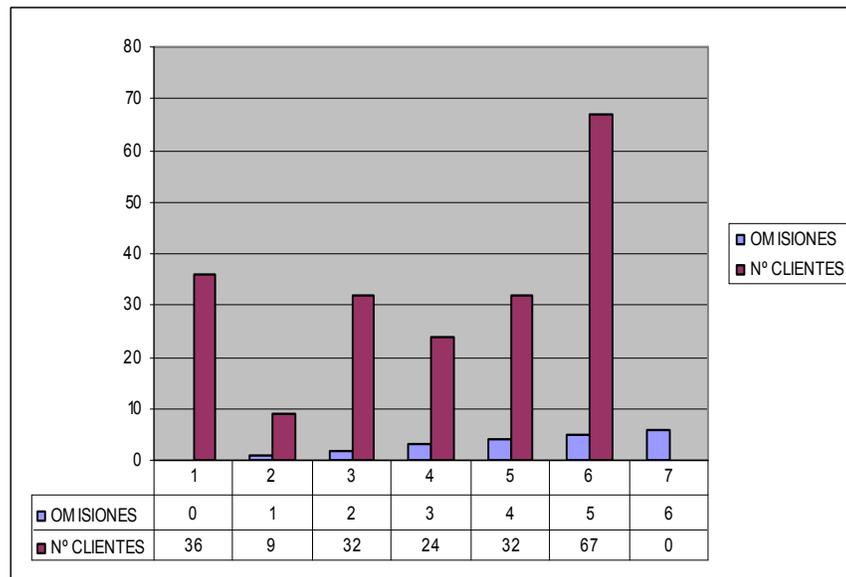
CAPÍTULO VI

PRUEBAS – EXPERIMENTOS NUMÉRICOS

6.1. Primera corrida del simulador convencional con la red neuronal

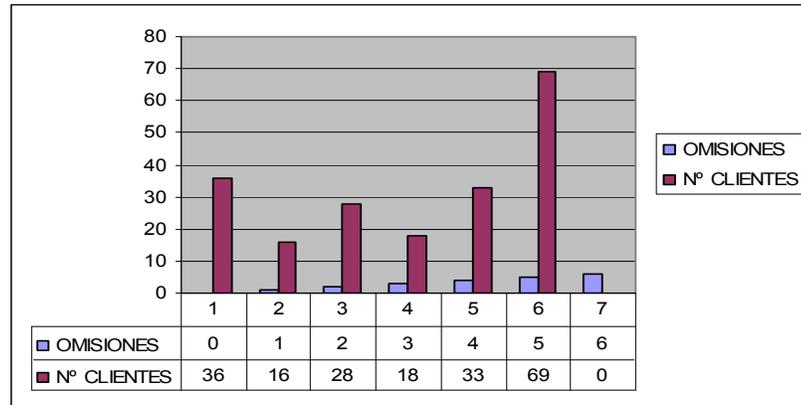
Una vez comprobado que la red neuronal ha aprendido, se efectúa una primera corrida del programa convencional con el RNA, para comprobar que el RNA es mejor PREDICTOR que el simulador convencional. Para eso se utiliza la prueba de Coeficiente de Correlación de Spearman y la técnica de Diferencia de Medias de Pearson. Se muestran las tablas (de 6.1.1 a 6.1.8) con sus cuadros (código de barras) correspondientes, tal como se dan a continuación:

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	36	2
1	9	6
2	32	3
3	24	5
4	32	4
5	67	1
6	0	7
	200	



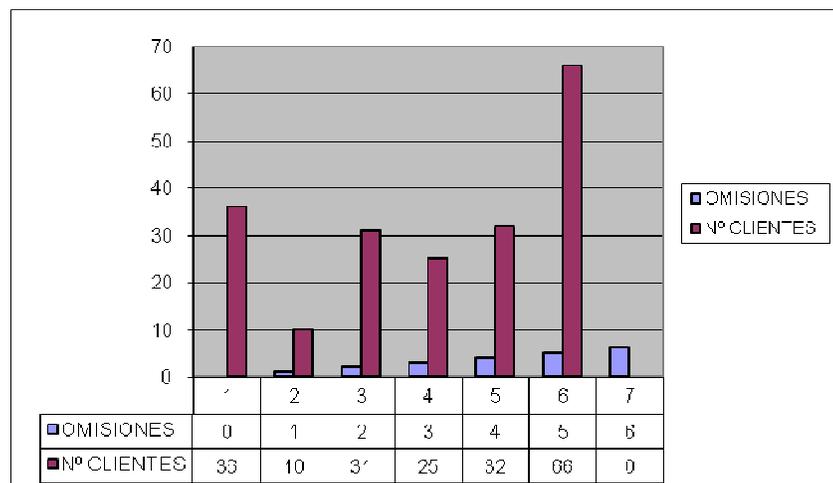
Cuadro 6.1.1. Datos Históricos **Reales** (6 meses anteriores)

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	36	2
1	16	6
2	28	4
3	18	5
4	33	3
5	69	1
6	0	7
	200	



Cuadro 6.1.2. Pronóstico del Programa **Convencional**
(6 meses próximos)

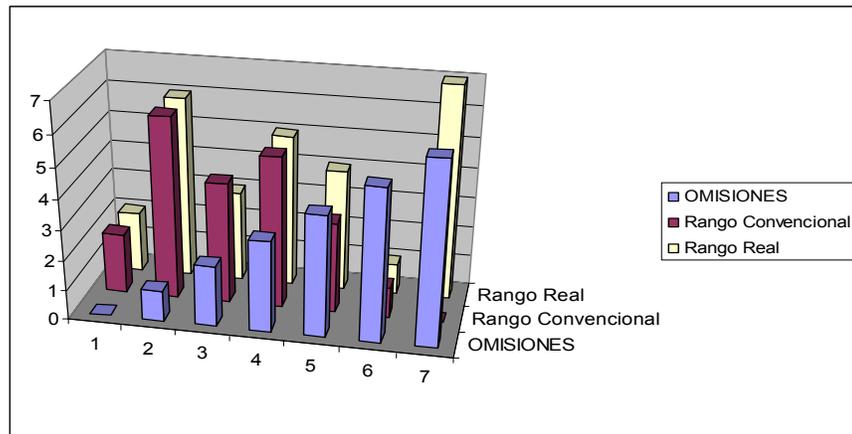
OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	36	2
1	10	6
2	31	3
3	25	5
4	32	4
5	66	1
6	0	7
	200	



Cuadro 6.1.3. Pronóstico del Programa **RNA** (6 meses próximos)

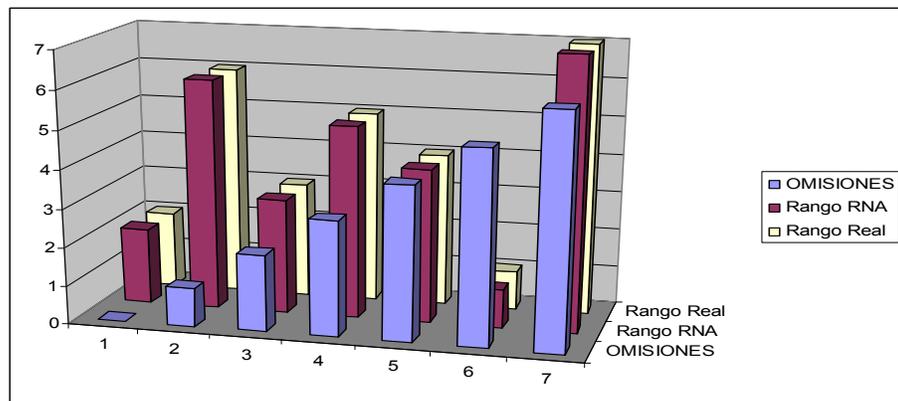
- Aplicando el **Coefficiente de correlación de Spearman**:

OMISIONES	Rango Convencional	Rango Real	D	D ²
0	2	2	0	0
1	6	6	0	0
2	4	3	1	1
3	5	5	0	0
4	3	4	-1	1
5	1	1	0	0
6	7	7	0	0
				0.96



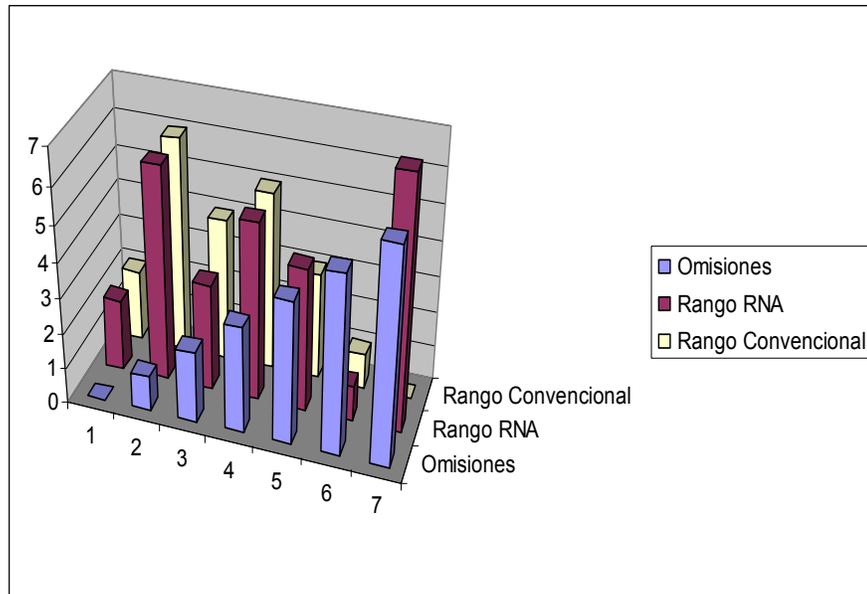
Cuadro 6.1.4. Correlación Convencional / Real

OMISIONES	Rango RNA	Rango Real	D	D ²
0	2	2	0	0
1	6	6	0	0
2	3	3	0	0
3	5	5	0	0
4	4	4	0	0
5	1	1	0	0
6	7	7	0	0
				1.00



Cuadro 6.1.5. Correlación RNA / Real

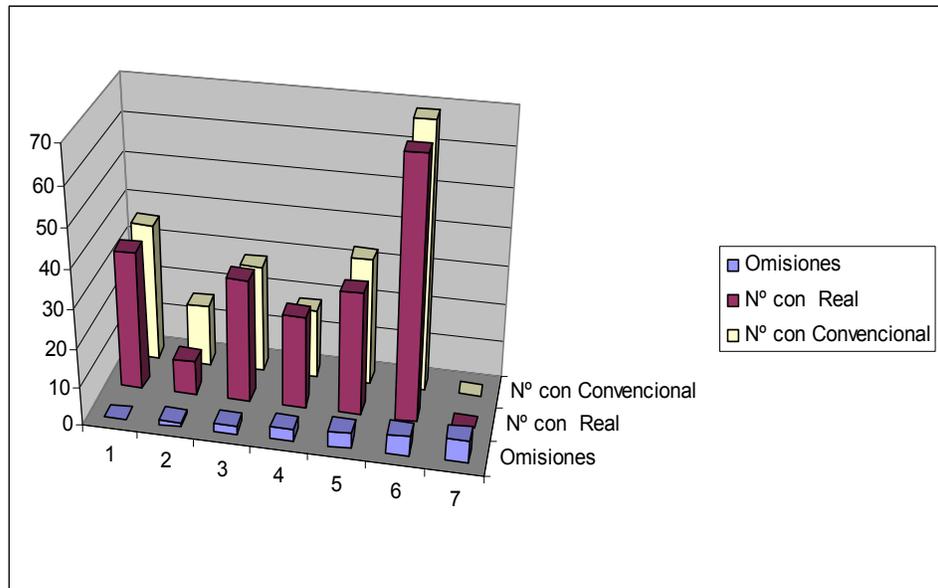
OMISIONES	Rango RNA	Rango Convencional	D	D ²
0	2	2	0	0
1	6	6	0	0
2	3	4	-1	1
3	5	5	0	0
4	4	3	1	1
5	1	1	0	0
6	7	7	0	0
				0.96



Cuadro 6.1.6. Correlación RNA / Convencional

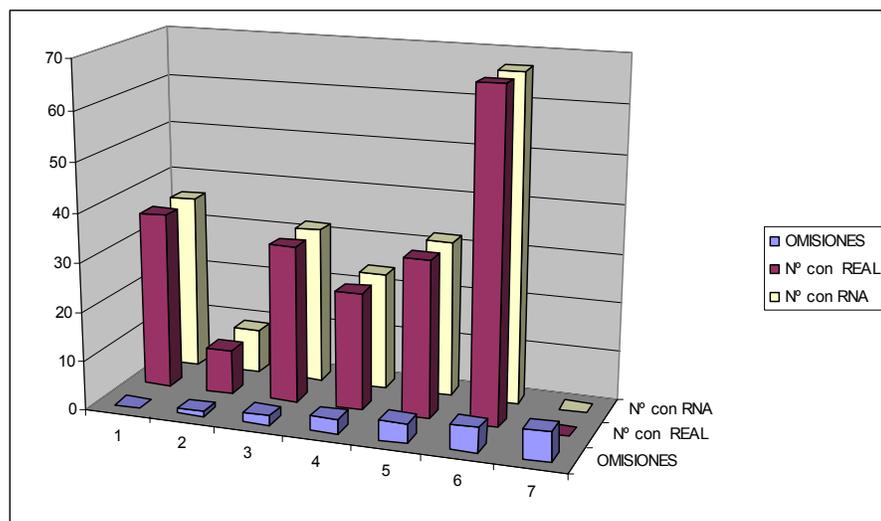
- Aplicando la prueba de **Independencia de la Diferencia de Medias de Pearson:**

OMISIONES	No. con REAL	No. con Convencional	Diferencia	Parámetros M1 y S1
0	36	36	0	
1	9	16	7	
2	32	28	4	Valor Medio
3	24	18	6	2.86
4	32	33	1	Desviación Estándar
5	67	69	2	2.85
6	0	0	0	
TOTALES	200	200		



Cuadro 6.1.7. Diferencias por tipo y cantidad de omisiones Real / Convencional

OMISIONES	No. con REAL	No. con RNA	Diferencia	Parámetros M2 y S2
0	36	36	0	
1	9	10	1	
2	32	31	1	Valor Medio
3	24	25	1	0.57
4	32	32	0	Desviación Estándar
5	67	66	1	0.53
6	0	0	0	
TOTALES	200	200		



Cuadro 6.1.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.

$$Z = \frac{M1 - M2}{\sqrt{(S1^2 / n1) + (S2^2 / n2)}}$$

Reemplazando los valores:

$$Z = \frac{2.86 - 0.57}{\sqrt{(2.85)^2 / 7) + (0.53)^2 / 7)} = 2.0830$$

Como el Z calculado es = 2,0830, es mayor al Z teórico (1,960).

A continuación se realizan cuatro (4) corridas adicionales de doscientos (200) clientes c/u en las secciones siguientes con la finalidad de comprobar que la RNA es mejor predictor que el Simulador Convencional.

6.2. Experimentos numéricos

Como parte de los experimentos numéricos se efectuarán cuatro (4) corridas adicionales para comprobar los resultados que el RNA es mejor predictor que el simulador convencional (Z teórico $> 1,960$).

6.2.1. Segunda corrida con la red entrenada

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	21	5
1	13	6
2	34	3
3	38	2
4	34	4
5	60	1
6	0	7
	200	

Cuadro 6.2.1.1. Datos Históricos **Reales** (6 meses anteriores)

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	22	5
1	8	6
2	38	3
3	41	2
4	35	4
5	56	1
6	0	7
	200	

Cuadro 6.2.1.2. Pronóstico del Programa **Convencional** (6 meses próximos)

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	21	5
1	12	6
2	36	3
3	38	2
4	32	4
5	60	1
6	1	7
	200	

Cuadro 6.2.1.3. Pronóstico del Programa **RNA** (6 meses próximos)

- Aplicando el **Coefficiente de correlación de Spearman**:

OMISIONES	Rango Convencional	Rango Real	D	D ²
0	5	5	0	0
1	6	6	0	0
2	3	3	0	0
3	2	2	0	0
4	4	4	0	0
5	1	1	0	0
6	7	7	0	0
				1.00

Cuadro 6.2.1.4. Correlación Convencional / Real

OMISIONES	Rango RNA	Rango Real	D	D ²
0	5	5	0	0
1	6	6	0	0
2	3	3	0	0
3	2	2	0	0
4	4	4	0	0
5	1	1	0	0
6	7	7	0	0
				1.00

Cuadro 6.2.1.5. Correlación RNA / Real

OMISIONES	Rango RNA	Rango Convencional	D	D ²
0	5	5	0	0
1	6	6	0	0
2	3	3	0	0
3	2	2	0	0
4	4	4	0	0
5	1	1	0	0
6	7	7	0	0
				1.00

Cuadro 6.2.1.6. Correlación RNA / Convencional

- Aplicando la prueba de Independencia de la **Diferencia de Medias de Pearson:**

OMISIONES	No. con REAL	No. con Convencional	Diferencia	Parámetros M1 y S1
0	21	22	1	
1	13	8	5	
2	34	38	4	Valor Medio
3	38	41	3	2.57
4	34	35	1	Desviación Estándar
5	60	56	4	1.90
6	0	0	0	
TOTALES	200	200		

Cuadro 6.2.1.7. Diferencias por tipo y cantidad de omisiones Real / Convencional

OMISIONES	No. con REAL	No. con RNA	Diferencia	Parámetros M2 y S2
0	21	21	0	
1	13	12	1	
2	34	36	2	Valor Medio
3	38	38	0	0.86
4	34	32	2	Desviación Estándar
5	60	60	0	0.90
6	0	1	1	
TOTALES	200	200		

Cuadro 6.2.1.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.

Reemplazando los valores:

$$Z = \frac{2.57 - 0.86}{\sqrt{(1.90)^2 / 7 + (0.90)^2 / 7}} = 2.1553$$

Como se ve el Z calculado = 2,1553, es mayor al Z teórico (1,960). Ello indica claramente la real Independencia de los métodos Real/RNA para esta tarea, como cabía esperarlo.

Cuadro 6.2.1.9. Datos de la segunda corrida

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Secuencia	Dni	Nombre	Orden	EntornoSc	Veracidad	ConductaEtic	Ahorro	AutoFinal	Ventajas	UnidadF	Respons	Conocimie	HistoriaL	CifraDeMerito
1	201	43754958	ESPINOZA PALACIOS JE	0.31	0.93	0.54	0.53	0.55	0.52	0.51	0.51	0.49	0.45	0.5	0
2	202	43348793	LOPEZ CONQUI RAFAEL	0.97	0.7	0.54	0.56	0.53	0.5	0.52	0.49	0.49	0.47	0.5	0
3	203	80084543	HUAMANI VARGAS WILL	0.6	0.27	0.25	0.24	0.25	0.19	0.22	0.22	0.2	0.16	0.17	0
4	204	21506974	CARBAJAL GUINTEROS	0.36	0.79	0.83	0.83	0.83	0.8	0.81	0.79	0.79	0.76	0.83	0
5	205	29525483	VILLEGAS CONDORI TEI	0.11	0.83	0.95	0.96	0.96	0.94	0.94	0.95	0.95	0.91	1	0
6	206	29538186	CHALCO GONZALES ZO	0.56	0.01	0.25	0.23	0.22	0.21	0.19	0.19	0.21	0.16	0.17	0
7	207	25526759	CABEZUDO DE CABELL	0.58	0.01	0.23	0.23	0.24	0.2	0.22	0.21	0.19	0.18	0.17	0
8	208	25536500	CARTOLIN ASTUPI#A AN	0.23	0.94	0.55	0.52	0.54	0.49	0.51	0.5	0.51	0.45	0.5	0
9	209	25539950	BALDASSARI AGUIRRE	0.32	1	0.7	0.69	0.7	0.66	0.66	0.65	0.64	0.62	0.67	0
10	210	25680329	VARGAS MINO PATRICIA	0.6	0.58	0.54	0.53	0.53	0.49	0.49	0.5	0.51	0.46	0.5	0
11	211	25683682	SOSA PAULINO MOISES	0.52	0.58	0.68	0.69	0.67	0.65	0.65	0.67	0.67	0.61	0.67	0
12	212	25683981	CANCINO FERNANDEZ E	0.09	0.07	0.22	0.23	0.23	0.21	0.2	0.21	0.19	0.16	0.17	0
13	213	25687798	CALDERON BARRIENTO	0.76	0.36	0.23	0.23	0.24	0.22	0.19	0.2	0.2	0.15	0.17	0
14	214	25694366	ARRIS PINEDA ENRIQUE	0.07	0.32	0.24	0.25	0.25	0.2	0.22	0.21	0.21	0.18	0.17	0
15	215	25695643	PALOMINO VDA DE GAI	0.76	0.58	0.95	0.97	0.96	0.94	0.94	0.94	0.94	0.92	1	0
16	215	25695643	PALOMINO VDA DE GAI	0.76	0.58	0.95	0.97	0.96	0.94	0.94	0.94	0.94	0.92	1	0
17	216	26600584	VIDAL VIDAL FEDERICO	0.41	0.23	0.69	0.69	0.67	0.66	0.64	0.64	0.64	0.62	0.67	0
18	217	26604624	APAZA NUPEZ SERAFIN	0.27	0.43	0.22	0.25	0.23	0.2	0.2	0.2	0.21	0.17	0.17	0
19	218	25617846	CANO ROJAS DE ROME	0.15	0.14	0.4	0.39	0.37	0.34	0.34	0.36	0.34	0.32	0.33	0
20	219	25620387	VERGARA ESPINOZA S,	0.28	0.08	0.55	0.54	0.53	0.49	0.49	0.49	0.5	0.48	0.5	0
21	220	25624156	BERNAOLA HEINSOHN I	0.13	0.42	0.38	0.39	0.38	0.34	0.34	0.34	0.35	0.33	0.33	0
22	221	25628351	PERICHE YENQUE NELI	0.66	0.67	0.39	0.38	0.4	0.37	0.35	0.34	0.36	0.31	0.33	0
23	222	25633900	MORALES DE CABELL	0.97	0.64	0.25	0.24	0.22	0.19	0.21	0.2	0.21	0.16	0.17	0
24	223	41443005	HUANA CAYLLAHUA RO	0.5	0.24	0.95	0.96	0.96	0.94	0.94	0.93	0.95	0.9	1	0
25	224	41656717	SANCHEZ SOTO AUGUS	0.2	0.76	0.97	0.97	0.95	0.94	0.93	0.93	0.93	0.91	1	0
26	225	41708644	HUANCA SANTOS ROMI	0.03	0.68	0.7	0.68	0.68	0.66	0.66	0.64	0.67	0.6	0.67	0
27	226	25650246	CHERO SILVA JOSE BE	0.14	0.28	0.25	0.23	0.23	0.19	0.2	0.19	0.19	0.18	0.17	0
28	227	25652316	SALDARRIAGA MARIN F	0.59	0.75	0.24	0.24	0.25	0.2	0.22	0.19	0.2	0.15	0.17	0
29	228	25655688	YATACO DE ARAUJO EL	0.72	0.58	0.23	0.22	0.25	0.19	0.2	0.19	0.21	0.17	0.17	0
30	229	41720773	VELEZMORO CORTEZ C	0.31	0.58	0.23	0.23	0.22	0.2	0.19	0.19	0.19	0.16	0.17	0
31	230	41727965	GASPAR CARDENAS RC	0.77	0.54	0.53	0.54	0.55	0.5	0.49	0.49	0.5	0.47	0.5	0
32	231	41777014	JAUREGUI PULGAR ALE	0.5	0.42	0.4	0.37	0.39	0.35	0.36	0.34	0.35	0.31	0.33	0
33	232	25666486	PASQUEL RIVAS PEDRI	0.18	0.1	0.97	0.96	0.97	0.93	0.93	0.93	0.95	0.91	1	0
34	233	41859323	TANTALIAN PEREZ JOSE	0.51	0.33	0.55	0.52	0.52	0.5	0.49	0.5	0.52	0.48	0.5	0
35	234	41909865	SALAZAR RODRIGUEZ N	0.4	0.42	0.25	0.23	0.23	0.22	0.21	0.19	0.2	0.18	0.17	0

6.2.2. Tercera corrida con red la entrenada

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	21	5
1	5	6
2	24	4
3	30	3
4	32	2
5	88	1
6	0	7
	200	

Cuadro 6.2.2.1. Datos Históricos **Reales** (6 meses anteriores)

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	22	5
1	7	6
2	23	4
3	26	3
4	34	2
5	88	1
6	0	7
	200	

Cuadro 6.2.2.2. Pronóstico del Programa **Convencional** (6 meses próximos)

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	21	5
1	6	6
2	23	4
3	30	3
4	32	2
5	88	1
6	0	7
	200	

Cuadro 6.2.2.3. Pronóstico del Programa **RNA** (6 meses próximos)

- Aplicando el **Coefficiente de correlación de Spearman**:

OMISIONES	Rango Convencional	Rango Real	D	D ²
0	5	5	0	0
1	6	6	0	0
2	4	4	0	0
3	3	3	0	0
4	2	2	0	0
5	1	1	0	0
6	7	7	0	0
				1.00

Cuadro 6.2.2.4. Correlación Convencional / Real

OMISIONES	Rango RNA	Rango Real	D	D ²
0	5	5	0	0
1	6	6	0	0
2	4	4	0	0
3	3	3	0	0
4	2	2	0	0
5	1	1	0	0
6	7	7	0	0
				1.00

Cuadro 6.2.2.5. Correlación RNA / Real

OMISIONES	Rango RNA	Rango Convencional	D	D ²
0	5	5	0	0
1	6	6	0	0
2	4	4	0	0
3	3	3	0	0
4	2	2	0	0
5	1	1	0	0
6	7	7	0	0
				1.00

Cuadro 6.2.2.6. Correlación RNA / Convencional

- Aplicando la prueba de Independencia de la **Diferencia de Medias de Pearson:**

OMISIONES	No. con REAL	No. con Convencional	Diferencia	Parámetros M1 y S1
0	21	22	1	
1	5	7	2	
2	24	23	1	Valor Medio
3	30	26	4	1.43
4	32	34	2	Desviación Estándar
5	88	88	0	1.40
6	0	0	0	
TOTALES	200	200		

Cuadro 6.2.2.7. Diferencias por tipo y cantidad de omisiones Real / Convencional

OMISIONES	No. con REAL	No. con RNA	Diferencia	Parámetros M2 y S2
0	21	21	0	
1	5	6	1	
2	24	23	1	Valor Medio
3	30	30	0	0.29
4	32	32	0	Desviación Estándar
5	88	88	0	0.49
6	0	0	0	
TOTALES	200	200		

Cuadro 6.2.2.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.

Reemplazando los valores:

$$Z = \frac{1.43 - 0.29}{\sqrt{(1.40)^2 / 7 + (0.49)^2 / 7}} = 2.0430$$

Como se ve el Z calculado = 2,0430, es mayor al Z teórico (1,960). Confirmando la independencia de ambos métodos.

Cuadro 6.2.2.9. Datos de la tercera corrida

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
	Secuencia	Idni	Nombre	Orden	Entorno	Veracidad	Conduct	Ahorro	AutoFin	Ventajas	Unidad	Fai	Respon	Conocir	Historial	CifraDe
				Ord	Ent	Ver	Con	Aho	Aut	Ven	Unid	Fai	Res	Con	His	Cif
2	401	4202026	BONILLA HURTADO	0.92	0.43	0.53	0.55	0.55	0.52	0.51	0.5	0.52	0.47	0.5	0	
3	402	7657799	BOJORQUEZ MENE	0.79	0.49	0.24	0.24	0.23	0.2	0.22	0.19	0.22	0.16	0.17	0	
4	403	25730332	CHAHUA HUAMANI	0.49	0.52	0.69	0.67	0.7	0.66	0.65	0.64	0.64	0.61	0.67	0	
5	404	1118685	TORRES RAMIREZ	0.96	0.47	0.55	0.54	0.55	0.5	0.51	0.51	0.5	0.45	0.5	0	
6	405	22409065	SANTAMARIA MAL	0.58	0.39	0.7	0.69	0.7	0.64	0.65	0.67	0.66	0.62	0.67	0	
7	406	22411258	CARRION ZEVALLC	0.8	0.86	0.22	0.23	0.25	0.19	0.21	0.2	0.19	0.17	0.17	0	
8	407	22512339	JUSTINIANO MARIA	0.4	0.07	0.96	0.96	0.97	0.95	0.93	0.95	0.94	0.91	1	0	
9	408	22512774	ESPINOZA TINEO T	0.09	0.6	0.39	0.38	0.39	0.35	0.34	0.34	0.35	0.33	0.33	0	
10	409	22521797	TRUJILLO ARGAND	0.6	0.85	0.24	0.25	0.23	0.2	0.2	0.19	0.19	0.17	0.17	0	
11	410	22702186	RAMOS ARIZA GUI	0.52	0.33	0.85	0.84	0.83	0.79	0.8	0.81	0.81	0.76	0.83	0	
12	411	22716615	CHAVEZ ASTETE J	0	0.16	0.23	0.23	0.23	0.22	0.22	0.21	0.21	0.18	0.17	0	
13	412	22717379	ROJAS VARGAS JU	0.76	0.8	0.53	0.54	0.53	0.51	0.49	0.49	0.49	0.48	0.5	0	
14	413	22993976	ESPINOZA AGUIER	0.13	0.71	0.82	0.84	0.82	0.81	0.8	0.82	0.8	0.77	0.83	0	
15	414	23265656	ACUÑA QUIISPE JE	0.88	0.25	0.37	0.39	0.39	0.35	0.35	0.35	0.37	0.32	0.33	0	
16	415	23268803	POMA HUALLPA E	0.16	0.1	0.22	0.23	0.24	0.22	0.2	0.19	0.22	0.16	0.17	0	
17	416	23467975	BALTAZAR LARA E	0.22	0.52	0.23	0.25	0.22	0.2	0.19	0.21	0.2	0.15	0.17	0	
18	417	23542993	NANEZ MALDONAC	0.06	0.7	0.54	0.54	0.54	0.5	0.51	0.5	0.49	0.45	0.5	0	
19	418	23549310	MORILLO HUARCA	0.06	0.25	0.69	0.68	0.69	0.65	0.64	0.64	0.67	0.62	0.67	0	
20	419	23679214	RAMOS AQUINO E	0.66	0.59	0.7	0.67	0.7	0.64	0.66	0.67	0.64	0.62	0.67	0	
21	420	23679305	MAMANI MAMANI E	0.79	0.85	0.95	0.95	0.97	0.93	0.93	0.95	0.93	0.92	1	0	
22	421	23693524	ASTO ROMAN MAF	0.94	0.08	0.84	0.82	0.84	0.8	0.81	0.81	0.79	0.78	0.83	0	
23	422	23816174	BENAVENTE ROM	0.5	0.93	0.25	0.23	0.23	0.21	0.21	0.2	0.2	0.15	0.17	0	
24	423	23893336	CUSIMAYTA GUIF	0.62	0.77	0.97	0.97	0.96	0.94	0.93	0.93	0.94	0.91	1	0	
25	424	23894040	NORIEGA GUZMAN	0.14	0.37	0.54	0.55	0.54	0.49	0.5	0.49	0.51	0.48	0.5	0	
26	425	23964804	SOTO HERRERA J	0.74	0.17	0.53	0.53	0.53	0.5	0.49	0.51	0.51	0.47	0.5	0	
27	426	24569612	ALCCALAICO CONI	0.84	0.84	0.69	0.68	0.7	0.67	0.69	0.66	0.66	0.61	0.67	0	
28	427	25416900	AQUIJE ARZOLA FI	0.84	0.34	0.24	0.23	0.24	0.21	0.19	0.21	0.22	0.18	0.17	0	
29	428	3820027	ANTON SANCHEZ S	0.54	0.91	0.38	0.39	0.4	0.36	0.37	0.34	0.37	0.3	0.33	0	
30	429	3824159	COCHRANE ARISM	0.85	0.65	0.22	0.24	0.25	0.21	0.22	0.2	0.19	0.17	0.17	0	
31	430	3834736	CHUICA OLAYA TE	0.05	0.78	0.25	0.23	0.23	0.19	0.2	0.2	0.19	0.16	0.17	0	
32	431	3839718	AGUIRRE ANZA GI	0.71	0.06	0.23	0.23	0.23	0.2	0.19	0.19	0.2	0.18	0.17	0	
33	432	3855286	RIVERA MEDINA S	0.83	0.12	0.23	0.23	0.24	0.2	0.22	0.2	0.21	0.18	0.17	0	
34	433	3856759	PENA OLIVOS RUE	0.01	0.21	0.97	0.97	0.96	0.93	0.93	0.93	0.93	0.92	1	0	
35	434	3861932	FARRO SUAREZ AI	0.78	0.45	0.38	0.38	0.38	0.35	0.37	0.34	0.34	0.33	0.33	0	

6.2.3. Cuarta corrida con la red entrenada

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	43	2
1	12	6
2	31	4
3	32	3
4	29	5
5	50	1
6	3	7
	200	

Cuadro 6.2.3.1. Datos Históricos **Reales** (6 meses anteriores)

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	52	1
1	7	6
2	33	3
3	29	4
4	28	5
5	51	2
6	0	7
	200	

Cuadro 6.2.3.2. Pronóstico del Programa **Convencional** (6 meses próximos)

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	42	2
1	11	6
2	32	4
3	33	3
4	28	5
5	52	1
6	2	7
	200	

Cuadro 6.2.3.3. Pronóstico del Programa **RNA** (6 meses próximos)

- Aplicando el **Coefficiente de correlación de Spearman**:

OMISIONES	Rango Convencional	Rango Real	D	D ²
0	1	2	-1	1
1	6	6	0	0
2	3	4	-1	1
3	4	3	1	1
4	5	5	0	0
5	2	1	1	1
6	7	7	0	0
				0.93

Cuadro 6.2.3.4. Correlación Convencional / Real

OMISIONES	Rango RNA	Rango Real	D	D ²
0	2	2	0	0
1	6	6	0	0
2	4	4	0	0
3	3	3	0	0
4	5	5	0	0
5	1	1	0	0
6	7	7	0	0
				1.00

Cuadro 6.2.3.5. Correlación RNA / Real

OMISIONES	Rango RNA	Rango Convencional	D	D ²
0	2	1	1	1
1	6	6	0	0
2	4	3	1	1
3	3	4	-1	1
4	5	5	0	0
5	1	2	-1	1
6	7	7	0	0
				0.93

Cuadro 6.2.3.6. Correlación RNA / Convencional

- Aplicando la prueba de Independencia de la **Diferencia de Medias de Pearson**:

OMISIONES	No. con REAL	No. con Convencional	Diferencia	Parámetros M1 y S1
0	43	52	9	
1	12	7	5	
2	31	33	2	Valor Medio
3	32	29	3	3.43
4	29	28	1	Desviación Estándar
5	50	51	1	2.82
6	3	0	3	
TOTALES	200	200		

Cuadro 6.2.3.7. Diferencias por tipo y cantidad de omisiones Real / Convencional

OMISIONES	No. con REAL	No. con RNA	Diferencia	Parámetros M2 y S2
0	43	42	1	
1	12	11	1	
2	31	32	1	Valor Medio
3	32	33	1	0.14
4	29	28	1	Desviación Estándar
5	50	52	2	0.38
6	3	2	1	
TOTALES	200	200		

Cuadro 6.2.3.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.

Reemplazando los valores:

$$Z = \frac{3.43 - 0.14}{\sqrt{(2.82)^2 / 7 + (0.38)^2 / 7}} = 2.1255$$

Como se ve el Z calculado = 2,1255, es mayor al Z teórico (1,960). Esto confirma la Independencia de ambos métodos.

Cuadro 6.2.3.9. Datos de la cuarta corrida

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Secuencia	Dni	Nombre	Orden	EntornoSc	Veracidad	Conducta	Ahorro	AutoFinan	Ventajas	UnidadFa	Respons:	Conocimr	HistorialDi	CifraDeMerito
2	601	16132020	ROSAS GOMEZ	0.33	0.74	0.37	0.38	0.37	0.36	0.36	0.35	0.34	0.33	0.33	0
3	602	25435995	CASTILLO GUEJ	0.51	0.26	0.23	0.23	0.23	0.21	0.19	0.22	0.19	0.17	0.17	0
4	603	7937416	A YONA LEON \	0.61	0.86	0.25	0.23	0.24	0.2	0.21	0.21	0.21	0.18	0.17	0
5	604	16586890	BALLONA TEZE	0.08	0.06	0.24	0.22	0.22	0.19	0.21	0.21	0.2	0.15	0.17	0
6	605	10043933	PAUCAR DAVA	0.02	0.02	0.4	0.38	0.38	0.34	0.35	0.35	0.35	0.32	0.33	0
7	606	29472840	ARNEZ GARCIA	0.48	0.64	0.24	0.24	0.24	0.2	0.2	0.2	0.2	0.16	0.17	0
8	607	18875818	MACO NIUNJA	0.35	0.09	0.95	0.97	0.97	0.93	0.94	0.95	0.94	0.91	1	0
9	608	8645621	LLAMOCA QUIE	0.16	0.57	0.97	0.97	0.95	0.94	0.93	0.93	0.93	0.91	1	0
10	609	8762836	MIENACHO MOI	0.65	0.72	0.97	0.96	0.96	0.94	0.93	0.93	0.93	0.91	1	0
11	610	9018674	HERNANDEZ R	0.59	0.87	0.97	0.96	0.96	0.94	0.94	0.94	0.94	0.9	1	0
12	611	7342809	PARI PAMPA S	0.18	0.78	0.39	0.4	0.4	0.36	0.35	0.37	0.34	0.33	0.33	0
13	612	7363510	MORON BULEJ	0.4	0.73	0.96	0.95	0.97	0.93	0.95	0.94	0.93	0.92	1	0
14	613	7371650	GARCIA GRILLC	0.16	0.86	0.39	0.37	0.38	0.35	0.35	0.34	0.36	0.31	0.33	0
15	614	7462322	LA COTERA CA	0.97	0.5	0.54	0.55	0.55	0.5	0.5	0.52	0.51	0.48	0.5	0
16	615	7484716	BARZOLA SANI	0.92	0.3	0.54	0.54	0.54	0.49	0.51	0.52	0.49	0.46	0.5	0
17	616	7537187	CARDENAS CA	0.74	0.79	0.83	0.85	0.84	0.79	0.8	0.79	0.78	0.78	0.83	0
18	617	7544839	PEÑA CORDOV	0.15	0.2	0.68	0.7	0.69	0.65	0.64	0.66	0.64	0.61	0.67	0
19	618	7556148	OGUINO OGUIN	0.4	0.36	0.96	0.96	0.97	0.93	0.93	0.94	0.93	0.9	1	0
20	619	7561440	APOLAYA MAN	0.9	0.13	0.23	0.23	0.25	0.2	0.2	0.19	0.21	0.18	0.17	0
21	620	7562478	OLMEDO RODR	0.55	0.86	0.55	0.55	0.55	0.51	0.5	0.5	0.49	0.48	0.5	0
22	621	7573825	BUSTAMANTE	0.65	0.02	0.24	0.26	0.26	0.2	0.21	0.19	0.19	0.18	0.17	0
23	622	7581798	VELASQUEZ O	0.85	0.2	0.39	0.38	0.4	0.36	0.35	0.35	0.34	0.32	0.33	0
24	623	7597283	REYNA GUEVA	0.31	0.12	0.68	0.69	0.7	0.67	0.65	0.64	0.67	0.62	0.67	0
25	624	7597370	MOLERO HERR	0.79	0.3	0.23	0.22	0.24	0.2	0.22	0.21	0.21	0.17	0.17	0
26	625	7598936	CARREÑO LE N	0.78	0.34	0.96	0.97	0.96	0.93	0.94	0.95	0.94	0.92	1	0
27	626	7599449	OLIDEN JARAM	0.72	0.98	0.97	0.97	0.97	0.93	0.95	0.93	0.94	0.91	1	0
28	627	7602763	GARCIA HIDALC	0.03	0.42	0.23	0.25	0.25	0.19	0.19	0.19	0.2	0.17	0.17	0
29	628	7610505	DAVILA ALVAN	0.56	0	0.52	0.53	0.54	0.5	0.51	0.52	0.52	0.46	0.5	0
30	629	7613023	VALLE MUNDA	0.11	0.86	0.23	0.25	0.24	0.2	0.2	0.22	0.19	0.15	0.17	0
31	630	7673052	REMY RONGUI	0.54	0.31	0.96	0.97	0.96	0.94	0.93	0.94	0.93	0.9	1	0
32	631	7689290	PALMA OLIVAR	0.75	0.46	0.7	0.7	0.68	0.65	0.67	0.64	0.67	0.62	0.67	0
33	632	7706289	PASCARELLA C	0.41	0.53	0.55	0.54	0.53	0.51	0.51	0.51	0.51	0.48	0.5	0
34	633	7708663	NUÑEZ PALACI	0.69	0.2	0.24	0.25	0.23	0.2	0.19	0.2	0.21	0.16	0.17	0
35	634	7731249	VALVERDE PR	0.62	0.97	0.85	0.85	0.83	0.81	0.79	0.79	0.8	0.78	0.83	0

6.2.4. Quinta corrida con la red entrenada

OMISIONES	Nº CLIENTES	ORDEN DE RANGO
0	13	5
1	9	6
2	16	4
3	26	2
4	24	3
5	112	1
6	0	7
	200	

Cuadro 6.2.4.1. Datos Históricos **Reales** (6 meses anteriores)

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	7	6
1	10	5
2	14	4
3	22	3
4	23	2
5	124	1
6	0	7
	200	

Cuadro 6.2.4.2. Pronóstico del Programa **Convencional** (6 meses próximos)

OMISIONES	N. CLIENTES	ORDEN DE RANGO
0	13	5
1	10	6
2	15	4
3	27	2
4	24	3
5	111	1
6	0	7
	200	

Cuadro 6.2.4.3. Pronóstico del Programa **RNA** (6 meses próximos)

- Aplicando el **Coefficiente de correlación de Spearman**:

OMISIONES	Rango Convencional	Rango Real	D	D ²
0	6	5	1	1
1	5	6	-1	1
2	4	4	0	0
3	3	2	1	1
4	2	3	-1	1
5	1	1	0	0
6	7	7	0	0
				0.93

Cuadro 6.2.4.4. Correlación Convencional / Real

OMISIONES	Rango RNA	Rango Real	D	D ²
0	5	5	0	0
1	6	6	0	0
2	4	4	0	0
3	2	2	0	0
4	3	3	0	0
5	1	1	0	0
6	7	7	0	0
				1.00

Cuadro 6.2.4.5. Correlación RNA / Real

OMISIONES	Rango RNA	Rango Convencional	D	D ²
0	5	6	-1	1
1	6	5	1	1
2	4	4	0	0
3	2	3	-1	1
4	3	2	1	1
5	1	1	0	0
6	7	7	0	0
				0.93

Cuadro 6.2.4.6. Correlación RNA / Convencional

- Aplicando la prueba de Independencia de la **Diferencia de Medias de Pearson:**

OMISIONES	N. con REAL	N. con Convencional	Diferencia	Parámetros M1 y S1
0	13	7	6	
1	9	10	1	
2	16	14	2	Valor Medio
3	26	22	4	3.71
4	24	23	1	Desviación Estándar
5	112	124	12	4.19
6	0	0	0	
TOTALES	200	200		

Cuadro 6.2.4.7. Diferencias por tipo y cantidad de omisiones Real / Convencional

OMISIONES	N. con REAL	N. con RNA	Diferencia	Parámetros M2 y S2
0	13	13	0	
1	9	10	1	
2	16	15	1	Valor Medio
3	26	27	1	0.57
4	24	24	0	Desviación Estándar
5	112	111	1	0.53
6	0	0	0	
TOTALES	200	200		

Cuadro 6.2.4.8. Diferencias por tipo y cantidad de Omisiones Real / RNA.

Reemplazando los valores:

$$Z = \frac{3.71 - 0.57}{\sqrt{(4.19)^2 / 7 + (0.53)^2 / 7}} = 1.9677$$

Como se ve el Z calculado = 1,9677, es mayor al Z teórico (1,960). Confirmando la independencia de los métodos

Cuadro 6.2.4.9. Datos de quinta corrida

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Secuencia	Dni	Nombre	Orden	Entorno	Verific	Conduct	Ahorro	AutoFinal	Ventajas	UnidadF	Respos	Conocimier	HistorialD	CifraDeM	Merito
2	801	21251582	TICLA DE ROJAS AI	0.45	0.54	0.68	0.68	0.68	0.64	0.65	0.66	0.65	0.6	0.67	0	
3	802	40642963	CARI RIOS MOISES	0.2	0.38	0.25	0.24	0.25	0.21	0.21	0.2	0.22	0.17	0.17	0	
4	803	19807615	ORREGO BEJARAN	0.82	0.25	0.25	0.23	0.25	0.19	0.19	0.2	0.21	0.18	0.17	0	
5	804	17822002	BURGOS ARGOME	0.85	0.99	0.39	0.37	0.37	0.34	0.35	0.35	0.34	0.3	0.33	0	
6	805	75641185	PUERTAS SALAZAR	0.69	0.61	0.84	0.85	0.84	0.8	0.82	0.81	0.8	0.77	0.83	0	
7	806	30763028	CERVANTES DIAZ	0.84	0.63	0.25	0.23	0.24	0.22	0.19	0.21	0.19	0.16	0.17	0	
8	807	10343480	ROMAN RODRIGUE	0.15	0.88	0.24	0.25	0.24	0.21	0.21	0.2	0.19	0.16	0.17	0	
9	808	6561805	MATOS ARIAS PAB	0.71	0.4	0.84	0.82	0.83	0.82	0.79	0.79	0.79	0.78	0.83	0	
10	809	17606756	VIGO ODAF SEGUN	0.89	0.79	0.25	0.25	0.23	0.19	0.19	0.19	0.19	0.18	0.17	0	
11	810	32777094	ROBLES GUERRER	0.76	0.3	0.22	0.22	0.24	0.2	0.21	0.2	0.2	0.16	0.17	0	
12	811	18897533	GALLARDO ALCALI	0.39	0.72	0.22	0.22	0.25	0.2	0.21	0.19	0.19	0.15	0.17	0	
13	812	29639684	LLERENA COSIO VI	0.66	0.52	0.82	0.84	0.84	0.81	0.79	0.79	0.8	0.75	0.83	0	
14	813	21819550	ALMEYDA DE SAR	0.24	0.09	0.23	0.22	0.23	0.19	0.21	0.21	0.2	0.18	0.17	0	
15	814	16579190	GUEVARA SANCHE	0.1	0.25	0.53	0.53	0.55	0.52	0.49	0.52	0.5	0.48	0.5	0	
16	815	16470982	CERCADO VASQUE	0.33	0.72	0.24	0.22	0.25	0.21	0.19	0.22	0.22	0.17	0.17	0	
17	816	16497374	RIVERA DE PERALI	0.34	0.07	0.25	0.25	0.23	0.22	0.21	0.22	0.19	0.17	0.17	0	
18	817	43360992	ZEVALLOS MELGAI	0.41	0.59	0.22	0.22	0.22	0.22	0.19	0.2	0.19	0.17	0.17	0	
19	818	30822968	QUIROZ DE BENAY	0.6	0.08	0.22	0.24	0.25	0.21	0.19	0.2	0.21	0.17	0.17	0	
20	819	29671770	ROBLES DE SOLIS	0.76	0.16	0.38	0.39	0.37	0.36	0.37	0.37	0.34	0.32	0.33	0	
21	820	25659593	AVILA OBLITAS CAI	0.81	0.27	0.25	0.24	0.23	0.19	0.19	0.19	0.19	0.18	0.17	0	
22	821	822562	RODRIGUEZ NAVAF	0.36	0.74	0.82	0.85	0.85	0.8	0.81	0.81	0.81	0.76	0.83	0	
23	822	40289406	LEON VILCHEZ JOS	0.18	0.7	0.55	0.54	0.54	0.5	0.5	0.52	0.5	0.45	0.5	0	
24	823	29289897	GALDOS VIZCARRA	0.09	0.61	0.4	0.39	0.4	0.36	0.35	0.36	0.36	0.3	0.33	0	
25	824	7962908	INFANTES LOPEZ G	0.64	0.19	0.22	0.24	0.25	0.2	0.21	0.19	0.19	0.17	0.17	0	
26	825	1124533	USHINAHUA PEREZ	0.7	0.97	0.54	0.55	0.54	0.49	0.49	0.52	0.49	0.46	0.5	0	
27	826	23823094	GUARNIZO CASTR	0.98	0.61	0.22	0.23	0.24	0.21	0.21	0.2	0.21	0.17	0.17	0	
28	827	1211107	CAHUI APAZA VALE	0.63	0.03	0.22	0.22	0.22	0.21	0.19	0.22	0.19	0.18	0.17	0	
29	828	7933910	CASTILLO CASTA#	0.92	0.28	0.24	0.25	0.25	0.19	0.22	0.22	0.22	0.17	0.17	0	
30	829	52269	SINARAHUA DEL AI	0.6	0.63	0.68	0.7	0.67	0.66	0.66	0.66	0.67	0.6	0.67	0	
31	830	17863612	ALVARADO CALDE	0.37	0.24	0.4	0.39	0.4	0.36	0.36	0.34	0.35	0.31	0.33	0	
32	831	41715712	GUIÑPE CACERES	0.28	0.99	0.23	0.25	0.24	0.19	0.22	0.21	0.19	0.15	0.17	0	
33	832	16573934	ZAPATA VDA DE AI	0.63	0.14	0.37	0.37	0.37	0.4	0.37	0.35	0.36	0.32	0.33	0	
34	833	43293307	CUNYA JORGE YIN	0.38	0.44	0.38	0.39	0.37	0.36	0.36	0.37	0.34	0.33	0.33	0	
35	834	42172439	FERREYRA ROMAN	0.27	0.31	0.24	0.23	0.24	0.2	0.19	0.2	0.19	0.17	0.17	0	

6.2.5. Resultados de los Experimentos – Convencional versus RNA

Como vemos en el cuadro siguiente en todos los experimentos comparativos el Z calculado fue mayor al Z teórico (1,960), lo cual demuestra que nuestra alternativa de cómputo es satisfactoria.

No. Experimento	Z calculado	> 1.960
Segunda	2.1553	SI
Tercera	2.0430	SI
Cuarta	2.1255	SI
Quinta	1.9677	SI

Figura 6.1. Resultado de los experimentos (Fuente: Autor)

Por lo tanto, con un 95% de confianza es preferible emplear el método del computo RNA que el Convencional en esta tarea específica de predecir la morosidad de los clientes, por ser mejor predictor y además cualitativamente más flexible y adaptable para esta tarea.

En nuestras pruebas hemos llegado entre el **97 y 99 % de exactitud o precisión** en las corridas efectuadas, lo cual demuestra que la RNA de tipo MLP, es muy eficiente como predictor para *Credit Scoring*.

CAPÍTULO VII

CONCLUSIONES Y RECOMENDACIONES

7.1. CONCLUSIONES

A continuación se indican las siguientes conclusiones obtenidas del presente trabajo:

- El Banco de la Nación podrá utilizar el modelo propuesto en sus aplicativos que involucren riesgos o incertidumbres como son: préstamos multired, préstamos de estudios, créditos hipotecarios, mejoras de vivienda, compra de deuda, etc.
- El modelo también permite al Banco de la Nación analizar otros tipos de modelos para otras áreas como son: Servicios Financieros, Tesorería, Riesgos, Fraudes, Seguridad de Información, Oficialía de Cumplimiento (lavado de activos), etc. con el apoyo de las RNA.
- Queda demostrado, dentro del margen de 95% de confianza estadística, que es preferible emplear el Método de cómputo **RNA** en lugar del **Convencional** para la tarea específica de predicción de morosidad para el nuevo servicio de Microcrédito, por ser **mejor predictor** del comportamiento de pagos futuro de los clientes.
- El tipo de software para predecir la morosidad con la **RNA**, es necesario y suficiente, para la gestión de un eventual servicio, con el uso de la herramienta **Easy NN Plus**, que demostró durante las pruebas realizadas, su superioridad.
- Los **criterios** de decisión para elegir el uso del software y/o hardware de cómputo neuronal óptimo para la tarea deben ser la **capacidad de pronóstico** prospectivo de los índices de riesgo de los futuros clientes y el **costo comparativo** de las soluciones neuronales disponibles.
- Finalmente, **el rendimiento operacional** en términos de **horas-hombre** de programación y cálculos, es marcadamente inferior en la solución de cómputo neuronal, lo que representa un factor adicional a tener en cuenta para la decisión de migrar a RNA.

7.2. RECOMENDACIONES

A continuación se indican las siguientes recomendaciones después de los resultados obtenidos del presente trabajo:

- Se sugiere el **estudio de otras variables** de los clientes de microcréditos de otras entidades de micro finanzas que se hayan realizado, así como **analizar el manejo de parámetros adicionales** que puedan afectar la decisión de aprobar o no un crédito a un cliente como son: asuntos legales, aspectos locales o regionales, etc.
- **Cuando** exista **data real** de Microcréditos en el Banco de la Nación, la red neuronal debe ser **retroalimentada**.
- Al Banco de la Nación, que continúe su modernización, en particular que adopte la tecnología de cómputo RNA para aquellos servicios nuevos que involucren riesgos o incertidumbres.
- A las entidades de Microcrédito, que adopten la tecnología de cómputo RNA en reemplazo de sus sistemas tradicionales. Esta decisión podría convertirlas en entidades competitivas y progresistas en este terreno.
- Al Colegio de Ingenieros del Perú, que establezca estándares de plataforma y de tipos de software de RNA para los proyectos públicos y privados. Ello fomentaría industrializar al país y convertirnos en exportadores de tecnología de cómputo.

REFERENCIAS BIBLIOGRAFICAS

Adel Lahsasna, Raja Noor Ainon, and Teh Ying Wah, Credit Scoring Models Using Soft Computing Methods: A Survey [Journal]. - Arabia: The International Arab Journal of Information Technology, 2010. - Vol. 7: Pg. 115.

Alistair Cockburn, Writing effective Use Cases [Libro]. - United States of America: Addison-Wesley, 2000.

Alsina, O. (coord.), La banca ética: mucho más que dinero [Libro].- Barcelona: FETS, 2001.

Audante Néstor, Credit Scoring para Pymes Financieras [Tesis].- Perú: Universidad Nacional de Ingeniería – FIIS, 2007.

Banco de la Nación – Portal, Plan estratégico 2009 - 2013 [En línea]. – Perú: Diciembre de 2011. <http://portal.bn.com.pe>.

Banco de la Nación – Revista mensual [Artículos]. – Perú: Febrero 2012.

Bernacki Mariusz y Wodarczyk Przemyslaw, Principles of training multi-layer neural network using backpropagation [En línea]. – September de 2004.

http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html.

Blum, Adam, Neural networks in C++: an objects-oriented framework for building connectionist systems [Libro]. – New York: John Wiley & Sons, 1992.

Braude Eric J., Ingeniería de Software - Una Perspectiva Orientada a Objetos [Libro]. - México: Editorial Alfaomega Editor, 2003.

Caudill, M. and Butler, Ch., Understanding Neural Networks: Computer Explorations [Libro]. - M.I.T. Press, 1992.

Cobos Carlos, Jhon Zuñiga, Juan Guarín, Elizabeth León y Martha Mendoza, CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos [Journal] – Colombia: Ingeniería e Investigación. Vol 30 No. 3, 2010.

Chapman Pete [y otros], CRISP-DM 1.0: Guide, Step by Step Data Mining [Libro]. - Londres, Nueva York y Bruselas: CRISP-DM Consortium, 2000. - Vol. 1: págs. 9-13.

Defu Zhang, Xiyue Zhou, Stephen C.H. Leung, Jiemin Zheng, Vertical Bagging Decision Trees model for Credit Scoring [journal]. – Holanda: Elsevier-Expert Systems with Applications, 2010. - Vol. 1.

Del Brio Bonifacio Martín y Sanz Molina Alfredo, Redes Neuronales y Sistemas Difusos [Libro]. - Madrid: Alfaomega-RA-MA, 2002. - Vol. 1: págs. 26-28.

Dieterlen, Paulette, La pobreza: un estudio filosófico [Libro]. – México: Instituto de investigaciones filosóficas - Fondo de Cultura Económica, 2003.

Faulín Javier, Ángel Juan. Simulación de Monte Carlo com Excel, Proyecto e-Math Financiado por la Secretaría de Estado de Educación y Universidades (MECD), 2005.

http://www.uoc.edu/in3/emath/docs/Simulacion_MC.pdf

Gallardo Arancibia, José, Metodología para el desarrollo de proyectos en Minería de Datos [Tesis]. 2009.

Hermosilla Raúl, Medición de Riesgos en la Evaluación de Crédito de las Microempresas [Libro]. - Perú. Editorial USMP – Curso de Marketing Estratégico, 2005.

Hernández Sampieri Roberto, Metodología de la Investigación [Libro]. – México: McGraw-Hill Ciudad de México, 1999.

Hertz John, Anders Krogh, Richard G. Palmer, Introduction to the theory of neural computation [Libro]. – California: Addison-Wesley, Redwood City, 1991.

Hilera José R. y Martínez Víctor J., Redes Neuronales Artificiales [Libro]. - Madrid: RAMA, 2000. - Vol. 1: Págs. 9-10.

Hussein Abdou, Genetic programming for Credit Scoring: The case of Egyptian public sector banks [journal]. – Holanda: Elsevier-Expert Systems with Applications, 2009. - Vol. 1.

Hussein Abdou, John Pointon, Ahmed El-Masry, Neural nets versus conventional techniques in Credit Scoring in Egyptian banking [journal]. – Holanda: Elsevier-Expert Systems with Applications, 2008. - Vol. 1.

Hussein Ali Hussein Abdou, Neural Nets and Genetic Programming versus Conventional Techniques [Tesis]. Plymouth Business School The University of Plymouth, 2009.

Isasi Viñuela Pedro y Inés M. Galván León, Redes Neuronales Artificiales un Enfoque Práctico [Libro]. - Madrid: Pearson Educación, 2004. - Vol. 1: Págs. 63-64.

Karlan Dean and Jonathan Zinman, Microcredit in Theory and Practice - Using Randomized Credit Scoring for Impact Evaluation [En línea]. – Science – Vol. 332, 2011. www.sciencemag.org

Kruchten Philippe, The Rational Unified Process [Libro] - United States of America: Addison Wesley, 2nd ed., 2000.

Martínez, 2000, Modelo no lineal en redes neuronales para clasificación. Una aplicación para determinar el riesgo en tarjetas de crédito. [Libro]. España - Departamento de Gestión y Métodos Cuantitativos, ETEA Córdoba, 2000.

Martens David, Bart Baesens, Tony Van Gestel y Jan Vanthienen, Comprehensible Credit Scoring models using rule extraction from support vector machines.[Journal] European Journal of Operational Research, 2007.

Mi Banco – Portal, Evolución del Microcrédito 2009 [En línea]. – Perú: Diciembre de 2009. <http://portal.MiBanco.com.pe>.

Mohammad Fereydon Kiani, Fariborz Mahmoudi. A New Hybrid Method for Credit Scoring Based on Clustering and Support Vector Machine [Journal]. – Iran: Islamic Azad University, Qazvin Branch, 2010.

Morales Vallejo Pedro, Estadística aplicada a las Ciencias Sociales. CORRELACIÓN Y COVARIANZA, Madrid: Universidad Pontificia de Comillas, 2008. edit@pub.upcomillas.es

Murray R. Spiegel, Teoría Elemental del muestreo, Teoría de la decisión estadística, Ensayos de hipótesis y significación, Schaum's Outline of Statistics (FE: 1961) [MRE: 2011]

<http://genealogy.math.ndsu.nodak.edu/id.php?id=5279>

Nnamdi I. Nwulu and Shola G. Oroja, A Comparison of Different Soft Computing Models for Credit Scoring [Journal] - World Academy of Science, Engineering and Technology 78, 2011.

Philip Joos, Koen Vanhoof, and Hubert Ooghe y Nathalie Sierens, Credit classification: a comparison of LOGIT models and Decision Trees, Gholamreza Nakhaeizadeh et al. (eds.), Application of machine learning and data mining in finance: European Conference on Machine Learning (ECML'98), Workshop notes; 24.April 1998, Págs. 59-70.

Ping Yao, Hybrid Classifier Using Neighborhood Rough Set and SVM for Credit Scoring [Journal]. – China: School of Economics & Management, Heilongjiang Institute of Science and Technology, Harbin, 2009.

Portocarrero Maisch Felipe, Microfinanzas en el Perú- Experiencias y perspectivas [Libro]. – Perú Universidad del Pacífico, 2003

http://www.cmseventos.com/presentaciones/perumicro_2009/cv/felipe_p_ortocarrero_maisch.pdf

Salvador Rayo, Juan Lara y David Camino, A Credit Scoring Model for Institutions of Microfinance under the Basel II Normative [Publication periodical]. - Scielo: Journal of Economics, Finance and Administrative Science, 2010. Vol. 15. - <http://www.scielo.org.pe>.

Sánchez Camperos Edgar y M.C. Alma Yolanda Alanis García, Redes Neuronales Conceptos fundamentales y aplicaciones a control automático [Libro]. - Madrid: Pearson Educación, 2006. - Vol. 1: Págs. 63-64.

SBS – Superintendencia de banca, Seguros y AFP - Perú [En línea]. – Diciembre de 2010. <http://www.sbs.gob.pe>.

Trivelli Carolina, Indigenous Peoples, Poverty and Human Development in Latin America [Libro] – Perú: G. Hall and H. Patrinos, editores. Nueva York: Palgrave Mcmillan, 2006.

Valcárcel Ascencios, Sergio, Sistema inteligente para calificar alpacas en función a su fibra fina [Tesis Maestría]. - Perú: Universidad Nacional Mayor San Marcos, 2011. - Vol. 1.

Yang Liu, The evaluation of classification models for Credit Scoring, Alemania: Institut für Wirtschaftsinformatik, Georg-August Universität Göttingen, 2002, Págs. 6 - 26.

Zan Huanga, Hsinchun Chen, Chia-Jung Hsu, Wun-Hwa Chenb, Soushan Wu, Credit rating analysis with support vector machines and neural networks: a market comparative study [journal]. – USA: Elsevier-BV, 2003.

ANEXOS

ANEXO A1

Entrevista a un experto en el Servicio de Microcrédito

- ¿Cuál es la capacidad de recuperación de un crédito en este servicio?
La capacidad de recuperación es del 97 %.
(Fuente: www.accion.org (Organización no gubernamental dedicada exclusivamente al microcrédito mundial))
- ¿Cuál es su opinión sobre los historiales de pago?
La mayoría de entidades de microcréditos no guardan historiales de pago, lo hacen a través de las centrales de riesgos.
- ¿Cuál es el grado de riesgo que se puede aceptar?
Es una variable del entorno macro, el grado de riesgo es del 6 % para las micro finanzas. Mantienen el riesgo hasta el final y no venden sus carteras de riesgo.
- ¿Cómo está cambiando la competencia con este tipo de riesgos?
Todos quieren ingresar a los microcréditos, debido a que ya no es de alto riesgo, más bien es de alta ganancia.
Existe mayor ganancia y el riesgo es mucho menor debido a que en este sector socioeconómico (D y E) se ha comprobado que su índice de retorno es del 97 %.
- ¿Se puede conseguir un listado de historial de pagos de seis (6) meses de una institución de Microcrédito?
Se pueden conseguir a través de las entidades reguladoras: Centrales de riesgos, INFOCORP, Informa Perú, etc. Si alguna institución cuenta con una herramienta no es lo suficientemente buena para evaluar el riesgo.

ANEXO A2

Entrevista a un Gerente Ejecutivo sobre el nuevo Servicio y su importancia crítica para el banco

1. Lucha contra la pobreza

El Perú está incrementando el rubro de las microfinanzas, como está sucediendo también en América Latina. Para el banco es un punto importante la lucha contra la pobreza como apoyo al estado. La problemática del banco es la tecnología, el personal y el riesgo que estaría disponible a aceptar.

El banco apoya en muchos sitios de extrema pobreza, muestra de ello se tiene el “Programa Juntos”, que apoya a los más necesitados a nivel nacional (470,000 beneficiarios, en su mayoría madres campesinas). Se podría iniciar un programa de “Haciendo negocio desde la pobreza” a su vez que se resuelven grandes problemas nacionales”.

2. ¿Qué iniciativas hay para el Microcrédito?

Actualmente, no es viable para el banco, por la gran cantidad de empresas dedicadas al Microcrédito. El banco no compite con ellos. Lo que si el banco está aportando con su estructura de red a nivel nacional.

Las empresas utilizan sus canales de atención para efectuar sus operaciones de Microcréditos como otro servicio más a través de convenios (comisiones por transacciones). Sobre todo en los lugares dónde estás empresas no están y el banco es la Única Oferta Bancaria (UOB), con cerca de 525 oficinas.

Esto si es viable en un futuro, dado que se da en varios países de América Latina, pero se deberá implementar en un nuevo modelo basado en la inclusión social, y bancarización, en el cual se considerará el dominio del proceso de microfinanzas dentro del dominio del proceso bancario tradicional.

3. ¿Cómo iniciar los Microcréditos a los pobres?

Se debe otorgar préstamos a personas que carecen de aval o propiedades. Se debe crear un método para que los pobres puedan

ser sujetos de crédito y saber si responderán a créditos pequeños basados en la confianza de grupos de la comunidad.

La forma de evaluación realizada por el agente de crédito especializado permitirá obtener los beneficios en mejores condiciones minimizando los riesgos.

4. ¿Qué perspectivas tiene la Gerencia del BN, respecto a su plan estratégico?

Tenemos que admitir que nos ha costado esfuerzo para que nuestros clientes tomen como suyo el lema de que somos “el banco que cambia la vida de los peruanos”.

Dentro de nuestro Plan Estratégico Institucional (PEI), la misión del BN es “Brindar soluciones financieras con calidad de atención, agregando valor, contribuyendo con la descentralización, ampliando nuestra cobertura de servicios, promoviendo la bancarización y la inclusión social”.

Justamente, la bancarización y la inclusión social van de la mano, porque es un proceso para incorporar a ciudadanos excluidos del sistema financiero bancario a fin de que puedan utilizar créditos bancarios, en zonas donde no exista oferta bancaria, así como en sectores no atendidos o de extrema pobreza.

5. Cantidad de capital a invertir en este servicio.

El banco promovió un fondo especial de S/. 200 millones de sus utilidades, para aquellas instituciones que ofrecieran menos tasas de interés a los microempresarios.

A la fecha se han colocado cerca de S/. 550 millones y cerca de 200 mil créditos a través de las PYMES en todo el país.

La alianza BN-IFI ha favorecido a miles de empresas, entregando S/. 294 millones; utilizados como capital de trabajo para emprender o consolidar su negocio. En el banco tienen el dinero e instrumentos suficientes para dar más créditos a los microempresarios.

Se podría crear un proyecto para implementar dos programas: uno de 6 millones de dólares en microcréditos y otro de 10 millones en negocios sociales, cantidades que podrían apoyar a luchar contra la pobreza extrema.

ANEXO A3

Análisis Documental de la conveniencia de este Servicio

Sobre la conveniencia de este servicio, podríamos dar como ejemplo, el cómo nuestra red puede apoyar:

Un microempresario (negocio de panadería y moto taxi), le resultaba muy caro y riesgoso viajar a la capital de su departamento (Chiclayo) desde su centro poblado “Tranca Sasape” – Túcume; para gestionar y obtener un crédito. Él tuvo su primer crédito con la **Edpyme “Alternativa”**, gracias a que la entidad de microfinanzas comparte algunas oficinas del banco en la región de Lambayeque, en este caso como UOB en Túcume.

El banco como Servicio de Corresponsalía tiene convenio con las siguientes Instituciones que otorgan Microcréditos:

- a) **Bancos:** Mi banco
- b) **CMAC:** Piura, Trujillo, Sullana, Paita, Ica, Pisco, Maynas, Huancayo, Tacna, Cuzco y Santa.
- c) **EDPYME:** Confianza, Crear, Alternativa, Solidaridad, Credivisión, Raíz y Edyficar.
- d) **CRAC:** Señor de Luren, Nor Perú, CREDINKA, Cajamarca y San Pedro.

ANEXO A4

Análisis de la Magnitud y Restricciones de este Servicio

El Banco de la Nación apoyando la bancarización e inclusión social tiene los siguientes proyectos:

- Desarrollar el plan de Inversiones para el fortalecimiento de agencias en zonas UOB considerando el compromiso para la inclusión social.
- Diseñar nuevos productos para zonas UOB en función de sus necesidades.

- Realizar/patrocinar cursos y seminarios para promover la cultura bancaria y financiera a nivel de Lima y Provincias (alfabetización financiera).
- Programa de Agilización de Inversiones para Gobiernos Locales y Regionales.

Las tareas que está realizando son las siguientes:

- Ampliar los Convenios con las IFI (Instituciones Financieras Intermediarias) dedicadas a las microfinanzas, como son las cajas municipales, cajas rurales, Edpymes, etc.
- Estudio del desarrollo económico, social y espacial para la incorporación de nuevos parámetros a considerar en el Crecimiento de las agencias a nivel nacional.
- Promover la suscripción de convenios con entidades públicas, principalmente en el ámbito regional (teniendo como entes motores a las propias agencias UOB).
- Evaluar las necesidades de servicios financieros de las poblaciones excluidas del sistema bancario, buscando su interrelación con los programas de desarrollo económico y social.
- Promover asociaciones público privadas para lanzar productos para sectores no bancarizados.

ANEXO B1

Actores de la Prueba



DESCRIPCION
Persona encargada de realizar las pruebas de simulaciones. Realiza las siguientes funcionalidades: <ul style="list-style-type: none">• Carga clientes y su historial de pagos• Realiza la simulación del pago convencional• Realiza la simulación de pago con la RNA• Compara los rendimientos de pago Convencional / RNA

ANEXO B2

Diagrama de Casos de Uso de las Pruebas



Casos de uso:

- CU1 = Cargar clientes y su Historial de pagos (carga por programa)
- CU2 = Simular Pago Convencional (desarrollado en Java)
- CU3 = Simular Pago con RNA (uso de paquete RNA)
- CU4 = Comparar Rendimientos Pago Convencional / RNA (análisis estadístico).

ANEXO B3

Especificación de Caso de Uso: Simular Pago Convencional

1. Breve descripción

El caso de uso permite al Investigador, simular los Pagos de los Clientes de forma convencional, utilizando el método de Montecarlo.

El sistema consultará los datos de los clientes con su historial de pagos y registrará las frecuencias de MC y las corridas efectuadas, además del pronóstico para los futuros seis (6) meses de pago.

2. Actor

Investigador

3. Flujo de Eventos

3.1. Flujo Básico

1. El caso de uso comienza cuando el Investigador invoca la ruta del programa de Simulación Convencional.
2. El sistema muestra la interfaz "Simulación Convencional – Monte Carlo" con los campos:
 - Parámetros de Simulación: Nro. de clientes y Nro. De corridas (simulaciones).
Además de las opciones: Iniciar Simulación, Limpiar, (Lista de Clientes (inhabilitado) y Tablas (inhabilitado).
3. El Investigador ingresa los parámetros de simulación.
4. El investigador solicita "Iniciar Simulación".
5. El sistema obtiene el historial de pagos de los clientes.
6. El sistema calcula las frecuencias, según las omisiones de pago de los clientes.
 - 6.1. El sistema agrupa las omisiones por frecuencias del [0-6] c/u de los clientes.
 - 6.2. El sistema calcula la probabilidad de cada una de las frecuencias de omisiones entre la suma total de los clientes.
 - 6.3. El sistema calcula la probabilidad acumulada.
 - 6.4. El sistema registra las frecuencias de las omisiones.
7. El sistema calcula las asignaciones.
 - 7.1. Se distribuye en forma ascendente los intervalos asociados a cada omisión (número aleatorio de inicio y fin) en base a la probabilidad acumulada.
 - 7.2. El sistema registra las asignaciones por omisiones.
8. El sistema ejecuta la "**Simulación de Montecarlo**", según la cantidad de corridas y la cantidad de clientes.
9. El sistema compara la corrida de mayor porcentaje de aciertos versus la suma de omisiones reales.
10. El sistema registra el resultado de la simulación convencional, validando el rango de la asignación de cada cliente.
11. El sistema actualiza a cada cliente sus omisiones simuladas y calcula la diferencia contra la omisión real.
12. El sistema muestra el MSG: "Operación realizada con éxito".
13. El sistema habilita las opciones "Lista de Clientes" y "Tablas".

14. Si el Investigador solicita la opción "Tablas"
 - 14.1. Ir al subflujo mostrar tablas.
15. Si el Investigador solicita la opción "Lista de Clientes"
 - 15.1. Ir al subflujo Listar Clientes.
16. El Investigador solicita "X", se cierra la interfaz y el caso de uso finaliza.

3.2. Sub Flujos

3.2.1. Listar Clientes

1. El sistema muestra la interfaz de búsqueda con los criterios de búsqueda: código y nombres.
2. El investigador ingresa el criterio de búsqueda.
3. El cliente solicita la opción "Buscar".
4. El sistema muestra la "Lista de clientes" con los siguientes datos: código, nombre, pagos de los seis (6) últimos meses, cantidad de omisiones, cantidad de **No omisiones**, diferencia y valor absoluto.
5. El investigador selecciona "X" se cierra la interfaz y finaliza el subflujo.

3.2.2. Mostrar Tablas

1. El sistema muestra la información de Frecuencias: omisiones, frecuencias, probabilidad y probabilidad acumulada.
2. El sistema muestra la información de Asignaciones: omisiones, Nro. Aleatorio inicial y Nro. Aleatorio final.
3. El investigador selecciona "X" se cierra la interfaz y finaliza el subflujo.

3.3. Flujos Alternativos

3.3.1. Simulación con error

Si la simulación detecta errores, se muestra el MSG "Existe un defecto en la operación" y finaliza el caso de uso.

3.3.2. Limpiar

Si el Investigador solicita la opción "Limpiar"

1. El sistema inicializa los datos de la corrida anterior: frecuencia, asignaciones y Simulación convencional.
2. El sistema inicializa las omisiones y diferencia de c/u clientes.

4. Requerimientos Especiales

1. Implementar correctamente el método de Montecarlo.
2. Haber calculado las omisiones reales por cada cliente.

5. Pre Condiciones

1. Disponible la lista de Clientes (historial de pagos).

6. Post Condiciones

1. En el sistema quedarán registrado las frecuencias, asignaciones y la simulación convencional.
2. En el sistema quedarán actualizados los clientes con sus omisiones reales, simuladas y la diferencia.

7. Prototipo

7.1. Ventana Inicial - Simulación Convencional



Operación realizada con éxito.

Simulación Convencional - MONTE CARLO

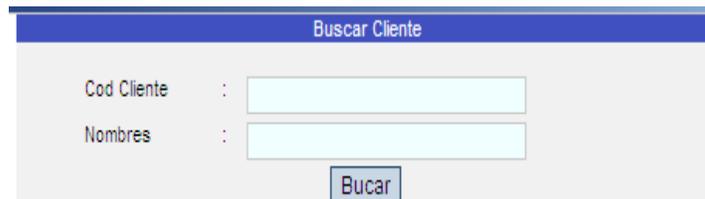
Número de clientes : 200

Número de corridas : 0.0

Iniciar Simulación Limpiar

Lista de Clientes Tablas

7.2. Ventana de búsqueda



Buscar Cliente

Cod Cliente :

Nombres :

Buscar

8. Otros

8.1. Información de Frecuencias.

Omisiones	Frecuencia	Prob.	Prob. Acu.
0	157	0.79	0.79
1	32	0.16	0.95
2	7	0.04	0.99
3	2	0.01	1
4	1	0.01	1
5	1	0.01	1
6	0	0	1

8.2. Información de Asignaciones.

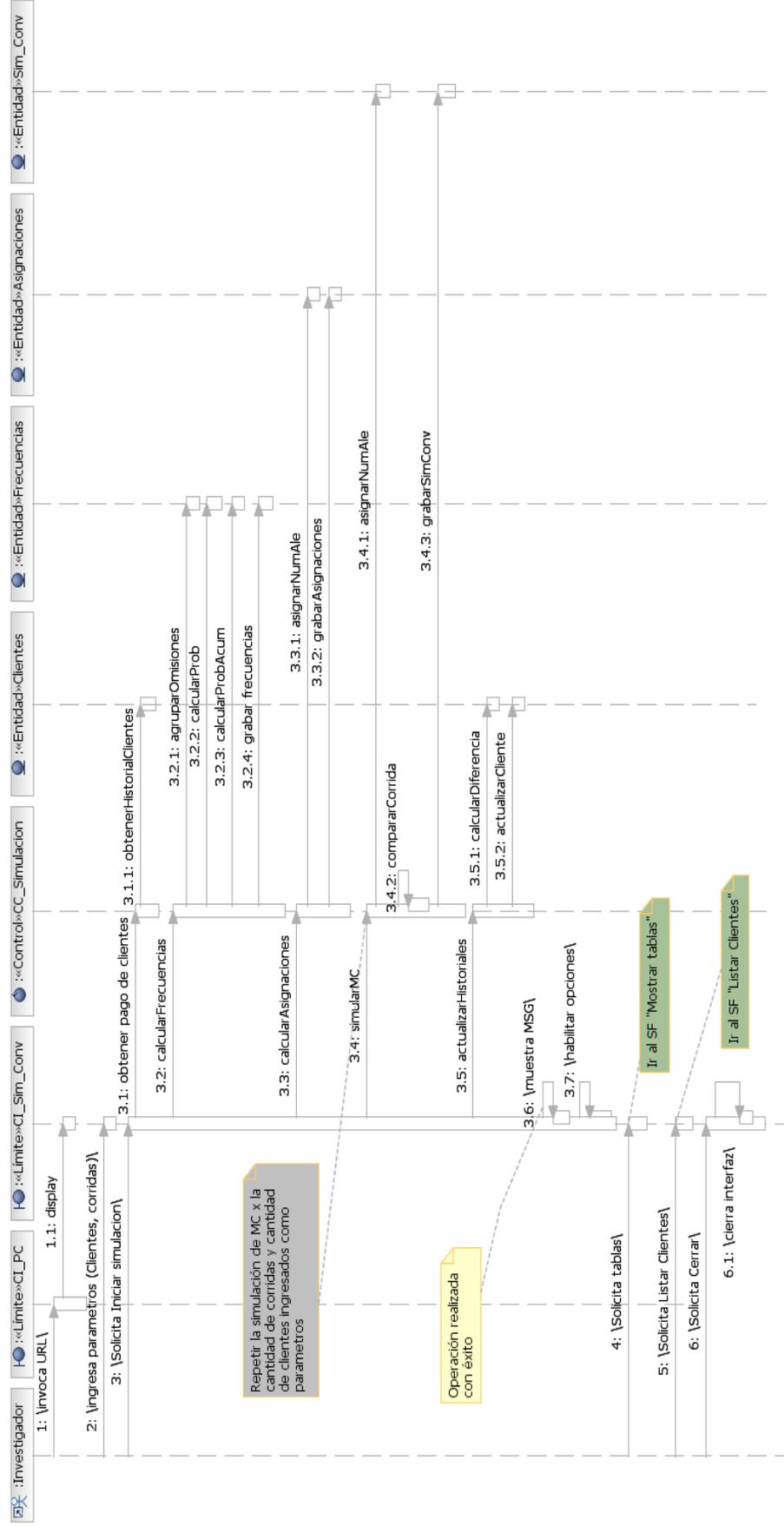
Omisiones	Aleatorio INI	Aleatorio FIN
0	0	78
1	79	94
2	95	98
3	99	99
4	100	100
5	100	100
6	100	100

8.3. Información de corrida Convencional.

Corrida	Código Cliente	Nombre	Nro Aleatorio	Omisión
32	207	RICHARD CHAVEZ AMEZQUITA	27	0
32	210	HERNAN CARDENAS TEJADA	86	1
32	215	RANULDO BOTTAZZI ALVAREZ	92	1
32	218	AGUSTIN MURGA LOPEZ	5	0
32	225	FELIX YEREN POMIANO	92	1
32	228	SEGUNDO PEREZ CONDOR	38	0
32	231	LUCINDA TABOADA ENCARNACION	73	0
32	234	LUIS CORREA BORJA	89	1
32	237	ELVI PACO ECHEVARRIA	11	0
32	240	ROBERTO SALAS BAELEA	24	0
32	243	MARIO SALVATIERRA HEREDIA	54	0
32	246	DIMAS ARRIETA ESPINOZA	98	2
32	249	FANNY MEJIA GARCIA	47	0
32	277	MIRTHA MARCOS OYOLA	51	0
32	294	AMELIDA PINTO CUADROS	87	1
32	354	ESTANISLAO HUAMAN ZAPANA	10	0
32	393	PABLO HUERTA LEON	68	0
32	401	NELLY ORMACHEA CABRERA	31	0
32	408	MARIO QQUECHO ORCCOHUARANCCA	5	0
32	413	LUIS CARBAJAL LUJAN	0	0
32	416	GABRIELA HERNANDEZ BACA	25	0

ANEXO B4

Diagrama de Secuencia



ANEXO B5

Diagrama de Clases Diseño (con operaciones)

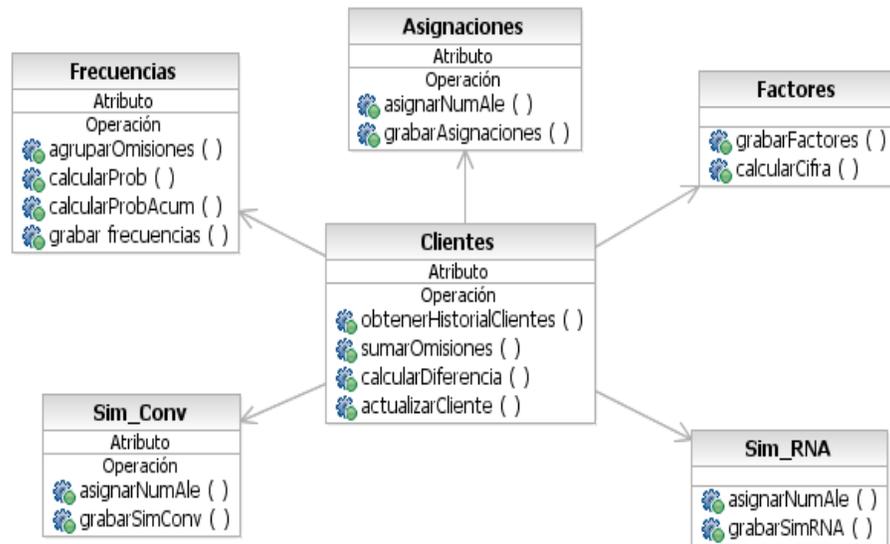


Diagrama de Clases de Diseño - Operaciones

ANEXO B6

ALGORITMO – Programa Convencional - Montecarlo

1. Breve descripción

El pseudocódigo permite correr el Algoritmo de Montecarlo y simular los Pagos de un Cliente de forma convencional, hasta completar los clientes (200) de Microcrédito de la base de datos con sus pronósticos de pago para los siguientes seis (6) meses.

La codificación se realiza como método en el programa java a desarrollar.

Antes de realizar las corridas se somete el código a “pruebas de escritorio” en tiempo real hasta estar seguros de no tener fallas de concepto o de sintaxis que puedan invalidarlo.

Finalmente se procede a insertar en el caso de Uso “Simular Pago Convencional”, para ser ejecutado con la opción “Iniciar Simulación”.

2. Previos

Para generar los 200 clientes de la Tabla Clientes se debe ejecutar cualquier programa estándar de generación aleatoria de nombres de personas y los 6 pagos (omisiones) se pueden extrapolar de los registros de entidades bancarias que ya estén dando este servicio y asignar aleatoriamente a cada cliente.

3. Pseudocódigo

```
Int Array 1[Tipo_Omision][Acumulador] = {0,0},{1,0},{2,0},{3,0},{4,0},{5,0},{6,0}.
Float Probabilidad_Acu = 0.
```

Con **CLIENTES**

Para cada **Cliente**

```
CLIENTES:: Suma_Omisiones ( ) // Calcula y registra el total de Omisiones del
                           Cliente
```

```
CLIENTES:: Frecuencia ( ) // Incrementa en 1 [Acumulador] del Array según
                           tipo de Omisión.
```

Fin_ Para clientes.

Con **FRECUENCIAS**

```
FRECUENCIAS:: Acumulador_ to_Frecuencia ( ). // Vuelca Array 1 a los 7
                                                registros, en campos Omisiones y Frecuencia respectivamente.
```

Con **FRECUENCIAS**

Para cada tipo de Omisión (7 registros)

```
FRECUENCIAS:: Frecuencia_ to_Probabilidad ( ) // Convierte cada Frecuencia
                                                en Probabilidad y registra este campo.
```

FRECUENCIAS:: Probabilidad_to_Probabilidad_Acu () // Suma Probabilidad en el Acumulador Probabilidad_Acu y registra total en este campo.

Fin_ Para Frecuencias.

Con **ASIGNACIONES**

Para cada tipo de Omisión (7 registros)

ASIGNACIONES:: Asignar () // Calcula número aleatorio inicial y final correspondiente. Registra estos valores en campos correspondientes.

Fin_ Para Asignaciones.

Con **CLIENTES** y **SIMULACION**

Para cada **Cliente**

SIMULACION:: Alea_Generado_to_Omisiones () // Genera número aleatorio con RAND (). Calcula número de omisiones que le corresponden y registra los campos.

Fin_ Para Clientes y Simulación.

Fin_Pseudocódigo.

ANEXO B7

Construcción de la Base de Datos

- a) Crear la tabla de **Cientes** con su historial de pagos para la prueba.
- b) Crear la tabla de **Frecuencias** de omisiones de pago mensual durante seis (6) meses.
- c) Crear la tabla de **Asignación** de intervalos de números aleatorios a las frecuencias.
- d) Crear la tabla de **Simulación** Convencional de omisiones por cliente.
- e) Crear la tabla de **Simulación RNA** de Omisiones por cliente.
- f) Crear la tabla de **Factores** de éxito del cliente.

Detalle de c/u de las tablas de la base de datos

- a) La tabla de Clientes tiene la siguiente estructura:

Atributos	Nulo	Dominio
Autocódigo	N	Clave autogenerada.
Nombre	N	Nombre y apellido – único.
Pago_mes_1	N	Booleano [Si = 1, No = 0]
Pago_mes_2	N	Ídem
Pago_mes_3	N	Ídem
Pago_mes_4	N	Ídem
Pago_mes_5	N	Ídem
Pago_mes_6	N	Ídem
Suma_Omisiones	S	Nro. omisiones reales [0 - 6]
Simul_Omisiones	S	Nro. omisiones simuladas [0 - 6]
Diferencia	S	Diferencia entre los 2 campos anteriores.

```
CODIGO SQL: CREATE TABLE Clientes
PRIMARY KEY Autocódigo
CREATE SECONDARY INDEX Nombre
ON Clientes.
```

- Esta Tabla identifica a los clientes y registra sus pagos mensuales **reales** hechos durante los últimos **seis meses** anteriores al mes de la prueba.
- No interesan los montos ni el orden en que aparecen estos pagos.
- El campo **Suma_Omisiones** registra la suma de los ceros de los seis campos de pago, es decir de las omisiones de pago acumuladas por el cliente en seis meses.

- El campo **Simul_Omisiones** registra el valor total de omisiones para los próximos seis meses **generado** por el algoritmo de Simulación Montecarlo para este Cliente.
- El campo **Diferencia** registra el valor **absoluto** de la resta entre los campos **Suma_Omisiones** y **Simul_Omisiones**.

b) La tabla de **Frecuencias** tiene la siguiente estructura:

Atributos	Nulo	Dominio
Omisiones	N	Número de omisiones de pago.
Frecuencia	S	Frecuencia de esta omisión.
Probabilidad	S	Probabilidad de ocurrencia.
Probabilidad_Acu	S	Probabilidad acumulada.

CODIGO SQL: **CREATE TABLE** Frecuencias
PRIMARY KEY Omisiones
CREATE SECONDARY INDEX Frecuencia
ON Frecuencias.

- Esta Tabla es un **cursor** que hace un resumen de la Tabla de **Cientes**, pagos de los clientes individuales en los 6 meses anteriores.
- Tiene 7 registros (uno para cada valor del campo **Omisiones** [0 a 6])
- Cada valor de **Omisiones** registra la suma de veces que aparece en la Tabla **Cientes** es decir su **Frecuencia**.
- La Frecuencia traducida como Probabilidad de aparición se registra en el campo **Probabilidad**.
- La suma progresiva y consecutiva (acumulativa) de las probabilidades de los siete registros es el valor de cada uno de los siete registros del campo **Probabilidad_Acu**.
- Notar que la suma de las 7 Probabilidades debe dar exactamente **1**, y el último registro del campo **Probabilidad_Acu** debe ser **1**.

c) La tabla de **Asignación** tiene la siguiente estructura:

Atributos	Nulo	Dominio
Omisiones	N	Número de omisiones de pago.
Nu_aleatorio_Ini	S	Nro. Inicial del intervalo
Nu_aleatorio_Fin	S	Nro. Final del intervalo.

CODIGO SQL: **CREATE TABLE** Asignación
PRIMARY KEY Omisiones
CREATE SECONDARY INDEX Nu_aleatorio_Ini
ON Asignación.

- Esta Tabla es un **cursor** con 7 registros que asigna un intervalo de números aleatorios consecutivos a cada uno de los 7 valores diferentes de omisiones (desde 0 en el primer registro hasta 6 en el último) teniendo en cuenta sus respectivas frecuencias. Cada intervalo está definido de manera única por su número aleatorio inicial y su número aleatorio final (campos **Nu_aleatorio_Ini** y **Nu_aleatorio_Fin** respectivamente).
- Los números aleatorios asignados son 100 y corren consecutivamente entre **00** y **99**.
- La Teoría de la Simulación Montecarlo enseña que cada intervalo tiene un número de números aleatorios proporcional a la probabilidad acumulada de cada frecuencia de las omisiones. Por ejemplo si el campo **Probabilidad_Acu** de uno de los 7 valores de Omisiones fuera **0,45** entonces debe asignársele como valor **Nu_aleatorio_Ini** el **45** y si el siguiente valor de Omisiones tuviese una probabilidad acumulada de **0,62** entonces el valor de **Nu_aleatorio_Fin** del registro de **Omisiones** que se está calculando será **61** y a todo número aleatorio generado por el Algoritmo que caiga entre **45** y **61** le corresponderá el mismo valor de **Omisiones**.
- Notar que para el cálculo del último registro (el séptimo) su **Nu_aleatorio_Fin** será **99** y el valor del **Nu_aleatorio_Ini** del primer registro será **00**.

d) La tabla de **Simulación Convencional** tiene la siguiente estructura:

Atributos	Nulo	Dominio
Codigo_corrida	N	Consecutivo generado – Entero.
Cliente	N	Código del cliente simulado.
Alea_generado	S	Nro. Aleatorio generado.
Omisiones	S	Número de omisiones del cliente próximo 6 meses.

CODIGO SQL: **CREATE TABLE** Simulación_Conv
PRIMARY KEY Código_Corrida
CREATE SECONDARY INDEX Cliente
ON Simulación_Conv.

- Esta Tabla registra consecutivamente las 999 corridas de simulación (una para cada cliente) para generar el número de omisiones de pago en que incurrirá este cliente en los próximos 6 meses después de la prueba.
- El número de omisiones lo calcula el Algoritmo a partir del valor del campo **Alea_Generado** (generado a su vez por la función RAND ()) entrando a la Tabla de **Asignación** y viendo cual es el valor de Omisiones que corresponde al intervalo donde cae el valor de **Alea_Generado**.
- El valor obtenido de Omisiones debe copiarse en el campo **Simul_Omisiones** de la Tabla de **Cientes**, en el registro correspondiente al cliente simulado.
- De esta manera la Tabla de **Cientes** queda completada para la Prueba Convencional.
- Se necesita una Tabla similar y adicional de **Simulación_RNA** que registre los valores de **Omisiones** calculados por RNA.
- Las pruebas estadísticas que comparan la calidad de las previsiones de ambos métodos de cálculo **solo necesitan** los valores de **Diferencias** de ambas Pruebas en la Tabla de Clientes. Con ellas, y solo con ellas, se realiza la Contrastación de Pearson.

e) La Tabla de **Simulación-RNA** tiene la siguiente estructura:

Atributos	Nulo	Dominio
Codigo_corrida	N	Consecutivo autogenerado – Entero
Cliente	N	Código del cliente simulado.
Cifra_Mérito	S	Cifra de los factores de éxito.
RNA_Omisiones	S	Número de omisiones del cliente próximo 6 meses.
RNA_diferencia	S	RNA_omisiones menos suma de omisiones.

CODIGO SQL: **CREATE TABLE** Simulación_RNA
PRIMARY KEY Código_Corrída
CREATE SECONDARY INDEX Cliente
ON Simulación_RNA.

- Esta Tabla registra consecutivamente las 999 corridas de simulación RNA (una para cada cliente) para generar el número de omisiones de pago en que incurrirá este cliente en los próximos 6 meses después de la prueba.
- El número de omisiones lo calcula y presenta en pantalla el Programa RNA a partir del valor del campo **Cifra_Mérito** (generado a su vez a partir de las calificaciones de sus factores de éxito) que se traduce en “pesos” internamente generados obtenidos inicialmente por medio del “aprendizaje” realizado previamente por el Programa RNA. El Programa genera el valor de Omisiones de cada cliente “pensando a la manera humana” con los factores de éxito del cliente.
- El valor obtenido de **RNA_Omisiones** debe restarse del valor **Suma_Omisiones** de la Tabla de **Cientes**, en el registro correspondiente al cliente simulado y calcular su valor de **RNA_Diferencia** para esta Tabla.
- **Cifra_Mérito** es un promedio ponderado de las notas obtenidas por el cliente al evaluar el banco sus factores personales de éxito.
- Con los valores de Diferencias de las dos pruebas, Convencional y RNA, ya pueden realizarse las pruebas Estadísticas de comprobación de las Hipótesis.

f) La Tabla de **Factores de Éxito** del cliente tiene la siguiente estructura:

Atributos	Nulo	Dominio
Cliente	N	Código del cliente simulado
Factor_1	S	Veracidad
Factor_2	S	Responsabilidad
Factor_3	S	Conducta ética
Factor_4	S	Entorno social
Factor_5	S	Unidad familiar
Factor_6	S	Orden

Factor_7	S	Conocimiento del negocio
Factor_8	S	Ventajas competitivas
Factor_9	S	Ahorro
Factor_10	S	Autofinanciamiento
Factor_11	S	Historial de pagos.
Cifra_Merito	S	Cifra actual media de los factores objetivos. Conversión a números de omisiones x el cliente.

CODIGO SQL: **CREATE TABLE** Factores_de_Exito
PRIMARY KEY Cliente
CREATE SECONDARY INDEX None
ON Factores_de_Exito.

- Esta Tabla registra consecutivamente las 999 x 11 calificaciones de 0 a 1 obtenidas por los clientes de la muestra al ser examinados por el Evaluador del Banco (un registro para cada cliente) y ser evaluados, desde el punto de vista del banco, sus factores de éxito en los negocios y por consiguiente su potencial de repago.
- Su importancia estriba en que el programa RNA, al ser entrenado, pondrá “pesos” a cada una de las 11 notas para obtener un Promedio Ponderado [De: 0 a 1] El Programa RNA “se dará cuenta” que los números de Omisiones son inversamente proporcionales a la Cifra de Mérito, pero la función exacta que los relaciona no es analítica ni expresable matemáticamente sino el resultado que emerge globalmente del trabajo de las “neuronas” del programa.

Los factores objetivos para evaluar un cliente son los siguientes:

- Veracidad** = (Nº de Documentos auténticos examinados) / (Nº total de Documentos examinados) [De: 0 a 1]
- Responsabilidad** = (Nº de Compromisos cumplidos) / (Nº total de Compromisos asumidos) [De: 0 a 1]
- Conducta ética** = (Nº de Clientes conformes) / (Nº total de Clientes entrevistados por el banco) [De: 0 a 1]
- Entorno social** = (Nº de Referencias favorables examinadas) / (Nº total de Referencias examinadas) [De: 0 a 1]

- e) **Unidad familiar** = (Nº de familiares que participan (conyugue vale 2)) / (Nº total de familiares directos) [De: 0 a 1]
- f) **Orden** = Nota del Evaluador después de visitar el negocio (Facilidad para ubicar documentos y general) [De: 0 a 1]
- g) **Conocimiento del negocio** = Nota del Evaluador con 5 preguntas (Precios, Mercados y Técnicas) [De: 0 a 1]
- h) **Ventajas competitivas** = Nota del Evaluador con 5 preguntas a clientes del negocio sobre satisfacción [De: 0 a 1]
- i) **Ahorro** = Nota del Evaluador después de 5 preguntas sobre las cuentas de ahorro que tiene el negocio [De: 0 a 1]
- j) **Autofinanciamiento** = Nota del Evaluador después de evaluar inversiones realizadas en el negocio. [De: 0 a 1]
- k) **Historial de pagos** = (Suma de pagos de los 6 últimos meses) / (Suma máxima de pagos posibles) [De: 0 a 1]

La fórmula que lleva al campo calculado **Cifra_Mérito** es la siguiente:

$$\text{Cifra_Mérito} = \frac{\sum P_k F_k}{\sum P_k} \quad (k = 1, 2, \dots, 11)$$

Donde:

F_k = k-ésima nota del k-ésimo factor

P_k = k-ésimo peso

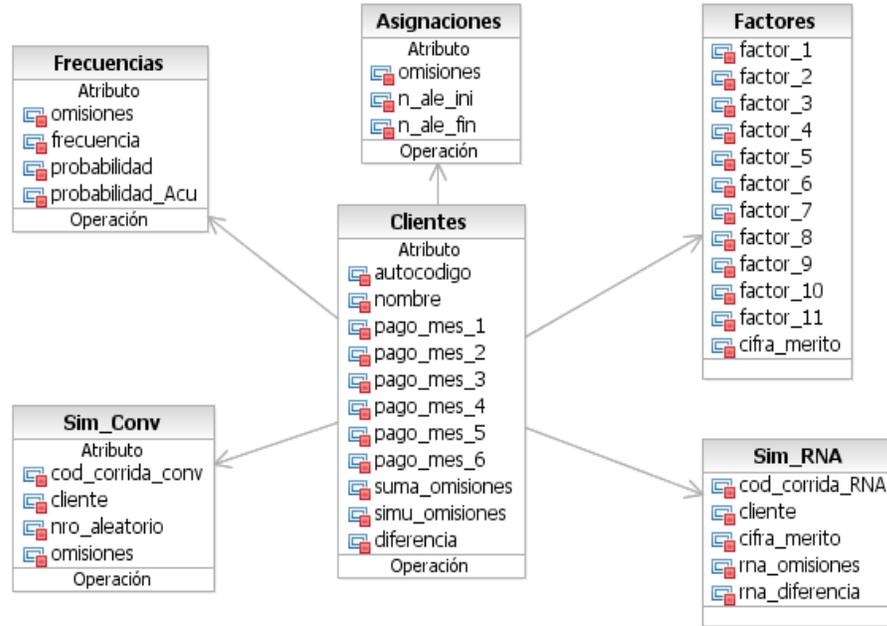
Para el caso de la comparación de la RNA con el Simulador Convencional el valor obtenido será el rango de omisiones (entre 0 y 6) que el cliente tendría.

Este es EL PARAMETRO FUNDAMENTAL del Programa RNA y lo realmente interesante es que prueba a los clientes NO BASADO EN DATOS HISTORICOS Y PROBABILIDADES como los programas convencionales sino “humanamente” apreciando el “perfil” de cada cliente según sus méritos personales.

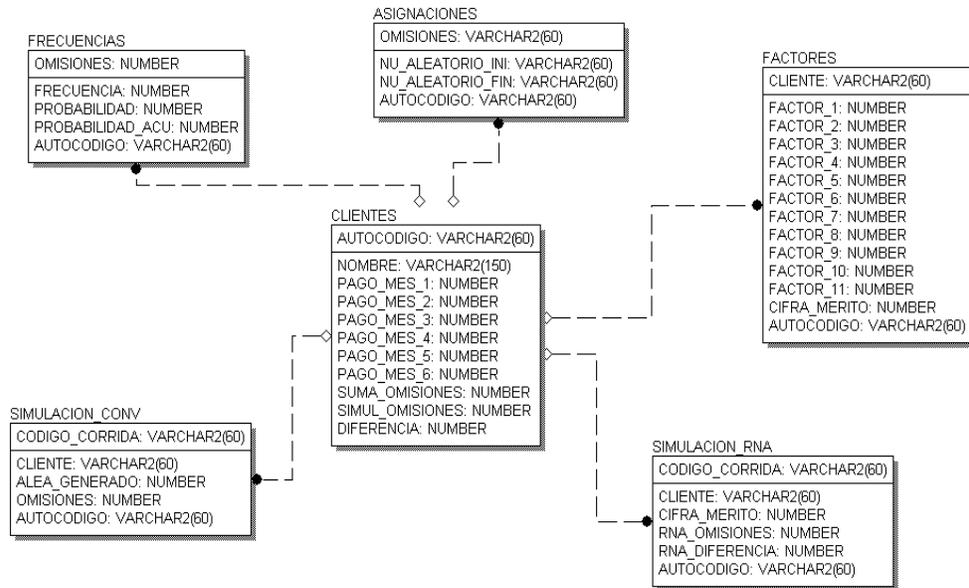
Otra función importante del RNA (que no tienen los programas convencionales) es que pueden mejorar su sensibilidad durante el aprendizaje variando los 11 pesos P_k y recalculando **Cifra_Mérito** hasta hallar aquel valor que más se acerque a generar un cuadro de

Omisiones igual al realmente registrado en los pasados 6 meses. De conseguirlo habrá superado al programa convencional y probado la hipótesis de este trabajo.

Creación de las Tablas



Anexo B7.1. Diagrama de Clases (con atributos)



Anexo B7.2. Esquema de Base de Datos

```

CREATE TABLE CLIENTES (
    AUTOCODIGO          VARCHAR2 (60) NOT NULL,
    NOMBRE              VARCHAR2 (150) NULL,
    PAGO_MES_1         NUMBER NULL,
    PAGO_MES_2         NUMBER NULL,
    PAGO_MES_3         NUMBER NULL,
    PAGO_MES_4         NUMBER NULL,
    PAGO_MES_5         NUMBER NULL,
    PAGO_MES_6         NUMBER NULL,
    SUMA_OMISIONES     NUMBER NULL,
    SIMUL_OMISIONES    NUMBER NULL,
    DIFERENCIA         NUMBER NULL,
    PRIMARY KEY (AUTOCODIGO)
);

CREATE TABLE FRECUENCIAS (
    OMISIONES          NUMBER NOT NULL,
    FRECUENCIA         NUMBER NULL,
    PROBABILIDAD       NUMBER NULL,
    PROBABILIDAD_ACU   NUMBER NULL,
    AUTOCODIGO         VARCHAR2 (60) NULL,
    PRIMARY KEY (OMISIONES),
    FOREIGN KEY (AUTOCODIGO)
        REFERENCES CLIENTES
        ON DELETE SET NULL
);

CREATE TABLE ASIGNACIONES (
    OMISIONES          VARCHAR2 (60) NOT NULL,
    NU_ALEATORIO_INI   VARCHAR2 (60) NULL,
    NU_ALEATORIO_FIN   VARCHAR2 (60) NULL,
    AUTOCODIGO         VARCHAR2 (60) NULL,
    PRIMARY KEY (OMISIONES),
    FOREIGN KEY (AUTOCODIGO)
        REFERENCES CLIENTES
        ON DELETE SET NULL
);

CREATE TABLE FACTORES (
    CLIENTE            VARCHAR2 (60) NOT NULL,
    FACTOR_1          NUMBER NULL,
    FACTOR_2          NUMBER NULL,
    FACTOR_3          NUMBER NULL,
    FACTOR_4          NUMBER NULL,
    FACTOR_5          NUMBER NULL,
    FACTOR_6          NUMBER NULL,
    FACTOR_7          NUMBER NULL,
    FACTOR_8          NUMBER NULL,
    FACTOR_9          NUMBER NULL,
    FACTOR_10         NUMBER NULL,
    FACTOR_11         NUMBER NULL,
    CIFRA_MERITO      NUMBER NULL,
    AUTOCODIGO        VARCHAR2 (60) NULL,
    PRIMARY KEY (CLIENTE),
    FOREIGN KEY (AUTOCODIGO)
        REFERENCES CLIENTES
        ON DELETE SET NULL
);

```

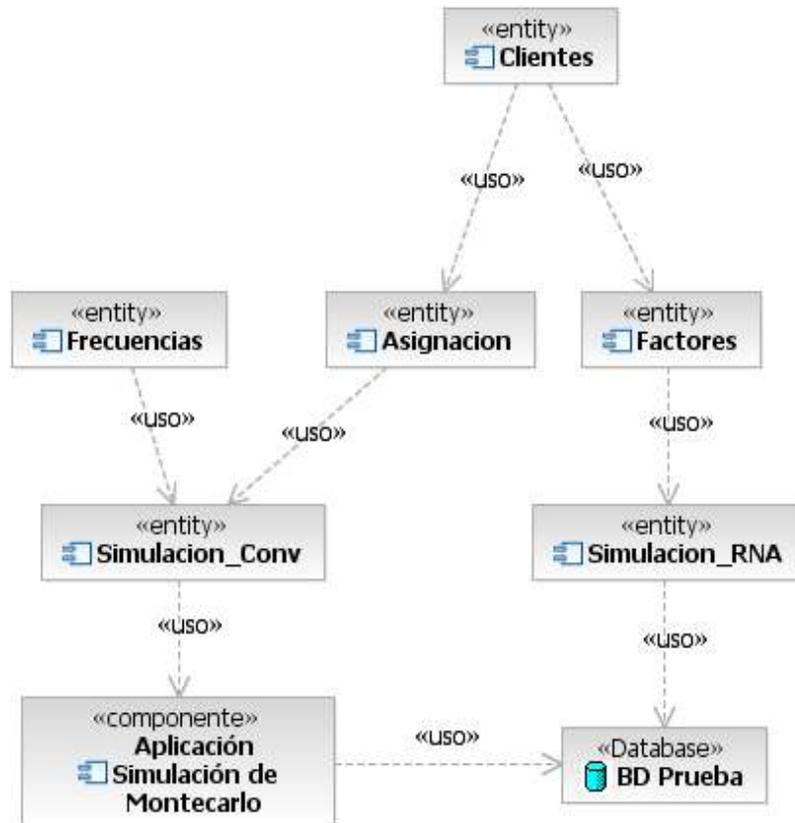
```

CREATE TABLE SIMULACION_CONV (
    CODIGO_CORRIDA      VARCHAR2 (60) NOT NULL,
    CLIENTE             VARCHAR2 (60) NULL,
    ALEA_GENERADO       NUMBER NULL,
    OMISIONES           NUMBER NULL,
    AUTOCODIGO          VARCHAR2 (60) NULL,
    PRIMARY KEY (CODIGO_CORRIDA),
    FOREIGN KEY (AUTOCODIGO)
        REFERENCES CLIENTES
        ON DELETE SET NULL
);
CREATE TABLE SIMULACION_RNA (
    CODIGO_CORRIDA      VARCHAR2 (60) NOT NULL,
    CLIENTE             VARCHAR2 (60) NULL,
    CIFRA_MERITO        NUMBER NULL,
    RNA_OMISIONES       NUMBER NULL,
    RNA_DIFERENCIA      NUMBER NULL,
    AUTOCODIGO          VARCHAR2 (60) NULL,
    PRIMARY KEY (CODIGO_CORRIDA),
    FOREIGN KEY (AUTOCODIGO)
        REFERENCES CLIENTES
        ON DELETE SET NULL
);

```

ANEXO B8

Diagrama de Componentes del Proceso

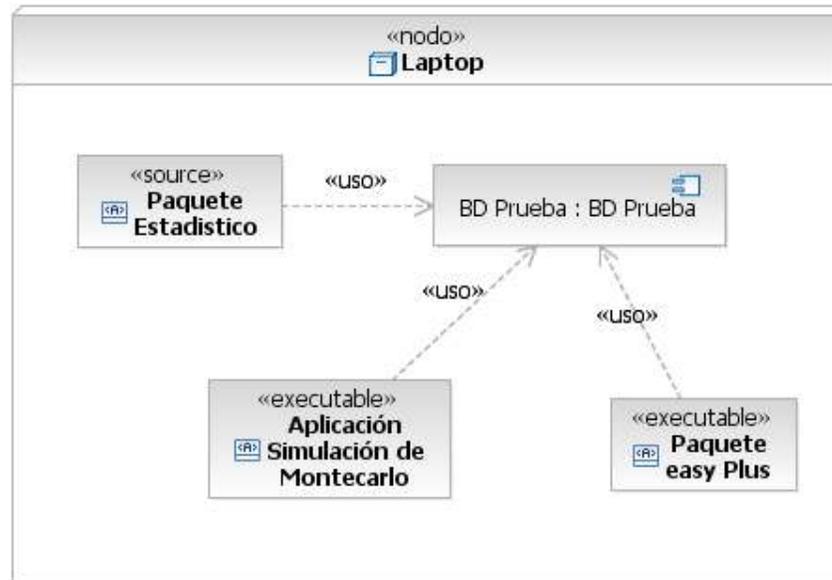


Componentes:

- **Aplicación** - Simulación Montecarlo.
- Tabla de **Cientes** de la Prueba (200 registros).
- Tabla de **Frecuencias** de Omisiones del pago de Clientes.
- Tabla de **Asignación** de intervalos de números aleatorios a Omisiones.
- Tabla de **Factores** de Éxito de Clientes.
- Tabla de **Simulación Convencional**.
- Tabla de **Simulación_RNA**.
- Base de Datos de la Prueba.

ANEXO B9

Diagrama de Despliegue



NODO Y COMPONENTES:

- Nodo – Laptop de Prueba.
- Base de Datos de la Prueba.
- Aplicación Convencional con Algoritmo de Simulación Montecarlo.
- Paquete RNA (EasyNN Plus).
- Paquete Estadístico.

ANEXO C

METODO DE MONTECARLO

Teniendo en cuenta lo dicho acerca de cómo funciona la Base de Datos durante el proceso de Pruebas comparativas, conviene establecer el fundamento de los algoritmos a emplear.

Al tratarse de una situación probabilística (ya que el hecho de pagar o no pagar es un fenómeno aleatorio, con o sin función de densidad de probabilidad que además, de existir, es totalmente desconocida), el **cómputo por Simulación Montecarlo** permite pronósticos individualizados para cada cliente, con el único inconveniente de necesitar frecuente actualización de las asignaciones de números aleatorios antes de cada corrida.

Procesos Aleatorios sin Ley de Probabilidad (Empíricas)

Este es el caso más frecuente en la práctica y por eso cuando se registra el historial de un proceso específico, como el caso del pago de deudas, se determina primero, por medio de la prueba estadística Chi-Cuadrado, si puede representarse por una Distribución teórica o no. Si no pasa la prueba (que es el caso más frecuente) entonces **no tendrá funciones de distribución teóricas** y estas son reemplazadas por tablas numéricas empíricas con los valores de la variable y su correspondiente probabilidad. Por ello son siempre discretas y su manejo se realiza por medio de la llamada **Simulación Montecarlo**.

Un ejemplo concreto es lo mejor para entenderlo y fijar ideas:

Supongamos que un banco necesita minimizar el riesgo de default (no pago) de préstamos, con el consiguiente menor costo por concepto del riesgo inherente.

El banco sigue la sencilla política de registrar el historial de pagos de todos los clientes durante 2 años (24 meses) registrando así los datos de la Tabla 1 para una muestra de 100 clientes tomados al azar:

N.Omisiones	4	5	6	7	8	9	10	11	12	13	14	15	16
Frecuencia	2	4	4	8	10	12	12	12	10	10	8	6	2
Probabilidad	0.02	0.04	0.04	0.08	0.10	0.12	0.12	0.12	0.10	0.10	0.08	0.06	0.02
Probabilidad Acumulada	0.02	0.06	0.10	0.18	0.28	0.40	0.52	0.64	0.74	0.84	0.92	0.98	1

Tabla 1 – Omisiones de pago mensuales de 100 clientes

Estos datos corresponden a 100 clientes y arrojan una media de 10,32 omisiones de pago en un lapso de 24 meses. Son clientes bastante malos sin duda pero ayudan a aclarar las ideas. Nótese que 4 y 16 son las omisiones mínima y máxima respectivamente. El riesgo es pues alto en este grupo.

Ahora, con esta Tabla, la **Simulación Montecarlo** sigue los siguientes pasos:

El primer paso es establecer el sorteo aleatorio con la siguiente tabla:

N. Omisiones	4	5	6	7	8	9	10	11	12	13	14	15	16
Intervalo de	00	02	06	10	18	28	40	52	64	74	84	92	98
Números	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Aleatorios VEB asignados	01	05	09	17	27	39	51	63	73	83	91	97	99

Tabla 2 – Sorteo de las Omisiones

Este proceso asigna **intervalos** de números VEB **a cada** Número de Omisiones registradas en la Tabla 1.

Debe notarse que la cantidad de números aleatorios VEB que se le asigna a cada valor de omisión es igual a la frecuencia de aparición de estos valores y como cada número aleatorio tiene la misma probabilidad de aparición eso significa que estos valores simulados aparecerán con la misma frecuencia que tienen registrada en el historial. También debe notarse que los números aleatorios VEB van de 00 a 99, o sea 100 números de 2 cifras cada uno en total.

El segundo paso del Método Montecarlo es establecer las series de corridas:

Por ejemplo tomemos una tirada o serie VEB de 7 números de 2 cifras

como:

47 – 17 – 22 – 94 – 67 – 03 – 38, uno para cada tipo de omisión

A ellos les corresponde la serie

10 – 7 – 8 – 15 – 12 – 5 – 9 de la Tabla 1 de N° de Omisiones

Ahora corremos una segunda serie VEB de 7 números:

73 – 97 – 81 – 31 – 00 – 39 – 25

A ellos les corresponde la serie

5 – 7 – 5 – 3 – 0 – 3 – 0 de la tabla de N° de Omisiones.

Ahora se establecen los parámetros de las corridas que se harán:

Número de clientes a simular = 100

Enseguida se corren las 6 x 100 tiradas y los resultados se vuelcan a una tabla de datos de 6 x 100 filas para la totalidad de los pronósticos conseguidos.

El tercer paso y último del Método Montecarlo es establecer los patrones mostrados por esta gran Tabla de Pronósticos que en este caso será compararlos con los 6 x 100 casos del método RNA. (Se usa la misma tabla después de inicializarla a cero pronósticos) Notar que en el caso RNA **se usan los factores de éxito** del cliente para “adivinar” si pagará o no este. No se usa el historial de pagos acumulado del Método Convencional ya que este es una técnica probabilística mientras que el RNA no lo es.

Conviene recordar que los campos PAGO_MES1 hasta PAGO_MES6 son los pagos **reales** de los clientes realizados **con anterioridad** a la prueba y por lo tanto sirven para crear la tabla 1 de la Simulación Montecarlo.