RESEARCH REPORT: MBL3

Credit Risk Analysis using Artificial Intelligence: Evidence from a Leading South African Banking Institution

Viresh Moonasar,

November 2007

Credit Risk Analysis using Artificial Intelligence: Evidence from a Leading South African Banking Institution

A research report presented to the

Graduate School of Business Leadership

University of South Africa

Submitted in partial fulfillment of the

requirements for the

MASTERS DEGREE IN BUSINESS LEADERSHIP,

UNIVERSITY OF SOUTH AFRICA

VIRESH MOONASAR

Student Number: 70792488

31 May 2007

Study Supervisor – Mr B S MASIYE

ABSTRACT

Credit risk analysis is an important topic in financial risk management. Financial institutions (e.g. commercial banks) that grant consumers credit need reliable models that can accurately detect and predict defaults. This research investigates the ability of artificial neural networks as a decision support system that can automatically detect and predict "bad" credit risks based on customers demographic, biographic and behavioural characteristics. The study focuses specifically on the learning vector quantization neural network algorithm.

This thesis contains a short overview of credit scoring models, an introduction to artificial neural networks and their applications and presents the performance evaluation results of a credit risk detection model based on learning vector quantization networks.

ACKNOWLEDGEMENTS

The work presented in this research report was carried out under the supervision of Mr Bryton Masiye. My gratitude goes to him for his assistance, contributions, coordination, supervision and guidance.

I also wish to thank:

- The financial institutions who provided the data for this research study.

- Finally, a special note of thanks to my family for their understanding, assistance and support.

Table of Contents

CHAPTER 3 - LITERATURE REVIEW

CHAPTER 4 - RESEARCH METHODOLOGY

CHAPTER 5 - RESEARCH RESULTS

List of Tables

List of Figures

List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
BNI	Bankruptcy Navigation Indicator
BNN	Back-propagation Neural network
BSS	Behavioural Scoring System
COP	Community of Property
CRF	Credit Research Foundation
CRG	Customer Risk Grade
DSS	Decision Support System
ETANN	Electrically Trainable Artificial Neural Network
FICO	Fair Isaac Corporation
GA	Genetic Algorithm
KNN	K-Nearest Neighbour
LMS	Least Mean Square
LSSFM	Least Squares Support Feature Machine
LVQ	Learning Vector Quantization
NCA	National Credit Act
MCLP	Multi-Criteria Linear Programming
MDA	Multiple Discriminant Analysis
MLP	Multi Layer Perceptron
NCA	National Credit Act
NFL	No Free Lunch
PC	Personal Computer
PCA	Principal Component Analysis
PI	Performance Indicator
RBF	Radial Basis Function
SME	Small Medium Enterprise
SOM	Self Organizing Map
SVM	Support Vector Machine

Chapter 1

Orientation

1.1. Introduction

Financial institutions (e.g. commercial banks) that grant consumers credit need reliable models that can accurately detect and predict defaults. One of the basic tasks which any finance institution must deal with, in the current competitive and turbulent business environment, is to minimize its credit risk. Scoring methods traditionally estimate the creditworthiness of a credit applicant or a current customer. They predict the probability that an applicant or existing borrower will default or become delinquent (Komor´ad, 2002).

Credit scoring is a key technique, in credit risk analysis, that assists organisations to decide whether or not to grant credit to consumers (Thomas, 2002). A common approach of credit scoring is to apply a classification technique on data of previous customers (both good credit customers and delinquent customers) in order to find a relationship between the customers characteristics and potential failure to service their debt. Institutions use credit scoring techniques (utilizing information from the consumers past credit history and current economic conditions) to determine which applicants will pay back their liabilities.

An accurate classifier is therefore essential to discriminate between new potential good and bad customers. This research report investigates the suitability and accuracy of artificial neural networks (ANNs) as a consumer credit risk classifier. ANNs have been previously applied to a diverse number of disciplines. This research project applies ANNs to the complex organisational problem of credit risk analysis in the South African financial management field. The research aims to test whether stochastic neural networks can improve the credit rating accuracy of conventional deterministic systems.

Demographic and biographic information, in addition to traditional financial and economic credit scoring information, is also used as inputs to the ANN classifier in order to determine if the ability and accuracy of predicting credit default can be increased.

1.2. Objectives of the research

The objective of this research report is to find reliable models that can improve the accuracy of predicting consumer credit defaults. Further, the significance and contribution of the research is to improve the quality of decision making by using ANNs as Decision Support Systems (DSS). The main objective of the study is to determine if the ANN system can improve the accuracy of existing credit rating models. This model is to be applied in conjunction with existing credit scoring systems. Institutions use credit scoring techniques (utilizing information from the consumers past credit history and current economic conditions) to determine which applicants will pay back their liabilities. However, there is still a percentage from the filtered successful applicants that do default i.e. all credit scoring systems have some degree of error. The objective of this research is to determine whether an ANN, applied in conjunction with the financial institution's expert systems, can be successful in reducing the number of consumer credit defaults in South Africa. The system will utilize the financial behavioural data in conjunction with biometric and demographic consumer information in attempt to increase the accuracy of credit default prediction.

Several specific ANN algorithms have been used for credit scoring in the past. The most popular algorithm used has been the back-propagation neural network (BNN). Komor´ad (2002) compared the results of multi-layer perceptron (MLP) and radial basis neural networks and found them to perform with similar results. The objective of this research is to evaluate the performance of the learning vector quantization (LVQ) neural network algorithm.

1.3. The research problem

The main research problem is to test whether LVQ artificial neural networks can be applied accurately to South African consumer credit risk analysis.

Sub-problems include the following:

a) Are there relationships between biographic and demographic characteristics of consumers and their credit risk scores?
b) Can LVQ ANNs successfully determine consumer credit rating classifications from the limited biographic, demographic and historic credit information available from banking institution databases?
c) How does the architecture and number of training cycles of the LVQ network affect the accuracy of the classification system?
d) How does the accuracy of the proposed LVQ network compare to existing classification systems currently used by the financial institutions?
e) How does the accuracy of the proposed LVQ network compare to actual consumer repayment behaviour? The actual repayment behaviour will be compared to the prediction of the ANN.

1.4. Definitions

Credit scoring and ANNs are defined in this section. These definitions are generally accepted in the academic literature and are applicable for the purpose of this research.

> Mester (1997:2) defines credit scoring as follows:
>
> "Credit scoring is a method of evaluating the credit risk of loan applications. Using historical data and statistical techniques, credit scoring tries to isolate the effects of various applicant characteristics on delinquencies and defaults. The method produces a "score" that a bank can use to rank its loan applicants or borrowers in terms of risk."

There is no single, universally accepted definition of a neural network. There are several definitions of neural networks that are generally accepted in literature. Although based ANNs are based on biological structure of the brain, Sarle (2002) emphasizes that biological neural networks are much more complex than the simple mathematical models used for **artificial** neural networks. Sarle (2002) describes a neural network as a network of many simple processors ("units"), each possibly having a small amount of local memory. The units are connected by communication channels ("connections") which usually carry numeric data, encoded by any of various means. The units operate only on their local data and on the inputs they receive via the connections.

According to Zurada (1992: xv):

“Artificial neural systems, or neural networks, are physical cellular systems which can acquire, store, and utilize experiential knowledge.”

Haykin (1994) defines neural networks aptly as

“A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network through a learning process.
2. Interneuron connection strengths known as synaptic weights are used to store the knowledge.”

1.5. Assumptions

a) There is a relationship between the biographic, demographic and credit history of consumers and their credit ratings as well as their actual ability to service their debt obligations.

b) Availability of data for training and testing of the ANN:
It is assumed that the data that has been acquired from one of South Africa’s leading banking institutions is accurate and of high integrity.

c) Suitability of the data:
The sample frame has been restricted to that supplied by the financial institution. The data supplied by the institution is the total credit card account population. Further, sampling is conducted in order to obtain training, validation and testing datasets for the ANN.

1.6. Delimitations

a) This study focuses on South African consumer credit ratings only. This study does not focus on the use of AI in company credit ratings or on credit data from outside of South Africa.

b) This study is constrained to LVQ ANN architectures for pattern classification. Other architectures, such as the popular back propagation algorithm, are not considered since previous literature suggests better accuracy with LVQs for pattern classification (Moonasar and Venayagamoorthy, 2001).

c) The study is further limited to individual models and not hybrid models of different AI technologies (Lai, Wang & Zhou, 2006a)

d) The dynamic aspects and structural changes over time are not considered in this study as the data to be obtained is not conducive to a longitudinal study (Lanquillon, 1999).

e) The study will not provide an explicit explanatory solution, if a relationship is found, between the consumer characteristics and their credit behaviour. It is assumed that the relationship is complex and contains multiple inter-dependencies. The objective is to predict who will default not to give explanations for why they default or answer hypothesis on the relationship between default and other economic or social variables (Thomas, 2002).

f) If hidden relationships are captured more effectively by Artificial Neural Network models compared to conventional discriminant analysis it in no way suggests substitution of experts by the software model. Rather the ANN model rating can be used as a bench mark indication to complement other methods e.g. financial expert/ consultant analysis. Therefore, an ANN model can successfully be used as a Decision Support System.

g) This study is conducted as simulations in software. Standard computers with the appropriate software are usually sufficient to run ANN applications. Specialized hardware can be used to run ANN programs more efficiently, but is not absolutely necessary. The 80170 chip developed by Intel and designated as an Electrically Trainable Artificial Neural Network (ETANN) is one example of ANN-specialized hardware.

1.7. Benefits of the Research Study

Credit Rating has become a crucial aspect for issuers and subscribers of debt. The practical utility of this study is in the fact that if a significant relationship is found with the help of an Artificial Neural Network, the same can be used as a benchmark from which expert/s can continue his/her analysis. This will considerably reduce the time and efforts of financial risk management experts because they can delegate routine analysis to the neural network model and concentrate on qualitative and subjective factors (Kumar and Haynes, 2003).

Other drivers affecting the need for businesses to adopt automated intelligent credit scoring are requirements placed on companies relative to compliance issues. As receivables represent such a significant portion of the asset base it is important to adequately state the true value of the portfolio by identifying its quality and the potential for bad debt write-off and corresponding reserves. It facilitates an accurate assessment of the true value of receivables as a liquid asset to be reflected on the balance sheet.

It is also inevitable that as the web begins to play a greater role in the order process, businesses will be required to adopt an automatic credit-scoring methodology to remain competitive in their ability to meet customer needs and expectations. Credit scoring provides clear benefits to the credit department; including speed, accuracy, consistency, reduction in bad debts, prioritization of collection activities and reduction in time required for risk assessment. Accuracy is assured because the review process is exempted from human error. Consistency is attained by using the same set of rules and weighted variables for review of the entire portfolio. Scoring permits regular review of the entire account base—quickly and efficiently identifying those accounts that require immediate attention for collection activity and isolating those customers that warrant further consideration through human intervention in the risk review process. The net effect is a substantial reduction in time required for risk assessment and a more systematic approach to the collection effort.

Neural network modelling is a statistical based scoring model. The basis for this type of model is a series of algorithms that are constantly and automatically being refined and updated with time and information pertinent to the model is gathered and incorporated into the model parameters. The advantage to this type of model is that it is not static and is constantly adjusting to changes, such as routine business and economic cycles that may have a positive or adverse effect on the outcome.

1.8. New contributions of this research report

Previous research on the use of ANNs has focused mainly on the BNN algorithm. This research focuses on LVQ classification algorithm.

Credit scoring techniques are normally based on previous financial history of the consumer. This work attempts to combine financial and economic factors with demographic and biographic consumer characteristics as inputs to the ANN classifier model.

This research is based on data received from South African banking institutions, and therefore tests national responsiveness of the ANN model to credit scoring.

1.9. Outline of the research report

This research report consists of six chapters. The List of References and Appendices sections follow the chapters, respectively. Apart from the last chapter, each of the chapters starts with an introduction section and concludes with a summary section.

Chapter 1: Orientation

This first chapter provides an introduction and a brief overview of the background, scope and context of the research topic. The research problem is introduced and the objectives of the research are discussed. In addition, definitions of credit scoring and ANNs are provided. This chapter also outlines the importance, assumptions, delimitations and the layout of the research report.

Chapter 2: Foundation of the research study

This chapter presents a theoretical foundation on credit scoring models and artificial neural networks. Part 1 of this chapter describes credit scoring models and the data utilized to calculate the consumer credit score. The problems associated with credit scores are also discussed. Part 2 deals with the description of ANN structures, architecture, and training. It concludes with unanswered problems facing ANNs.

Chapter 3: Literature review

Chapter Three covers the review of relevant theories and literature that were used in the research such as academic papers, journals, reports, the Internet, etc. The main sections of the chapter are the credit scoring principles and neural network approaches to consumer risk analysis. Each of these main sections has its sub-sections providing deep and detailed discussion of the literature.

Chapter 4: Research methodology

Chapter Four presents the research methodology and design strategy employed by this research work. The system model, data analysis techniques, sampling methods and limitations of this study are described in this chapter.

Chapter 5: Research results

The research results and interpretation of the results are presented in chapter 5. This chapter focuses on the presentation and statistical analysis of the data. The first section covers the preparation of data for analysis, the exploration and presentation of data using statistics. This section also explores the relationships and strengths of association between consumer characteristics and credit risk using statistical methods.

Chapter 6: Discussions, conclusions and recommendations

Chapter six contains a discussion of the overall research and the results of the research analysis of chapter five. Recommendations, based on the results of the analysis, are also presented in this chapter. Possible future directions of research in this field are also proposed. Finally, concluding remarks are made, explaining how the research objectives were achieved.

1.10. Summary

Financial institutions need reliable models that can accurately predict consumer defaults in order to minimize its credit risk. Credit scoring is a key technique that assists organisations to decide whether or not to grant credit to consumers.

ANNs are defined as simplistic models of biological neural networks that learn through experience. The objective of this research report is to find reliable models, based on ANNs that can improve the accuracy of detecting and predicting consumer credit defaults. Further, the significance and contribution of the research is to improve the quality of decision making by using ANNs as Decision Support Systems (DSS). This work is based on the LVQ neural network algorithm and is assessed on data obtained from the South African financial sector.

Chapter two, which follows this chapter, provides a theoretical grounding and foundation of credit scoring and artificial neural networks for this research.

Chapter 2

Foundation of the study

2.1. Introduction

This chapter provides the theoretical foundation of this study. This chapter is divided into two parts. The first part describes credit scoring models for financial risk analysis and the second part reviews basic neural network structures, theories and algorithms.

2.2. Credit Scoring Models

A credit score of a person is a numerical expression representing the creditworthiness of that individual, which is the likelihood that the individual will pay his or her debts (Wikipedia, 2007a).

Lenders, such as banks, use credit scores to evaluate the potential risk posed by lending money to consumers and to mitigate losses due to bad debt. Credit scores can be used to determine who qualifies for a loan and at what credit limits.

The Fair Isaac Corporation (FICO) provides the most well-known and widely used credit score model in the United States. The FICO score is calculated by applying statistical methods to ones credit information. Banks use credit scores as a factor in their lending decisions and may deny credit or charge higher interest rates if the score is low. FICO scores show how likely it is that a borrower will default. A separate score, bankruptcy navigator index (BNI), is used to indicate likelihood of bankruptcy.

Eurofinas, the European Federation of Finance House Associations, indicates that new consumer credit and outstanding consumer credit growth has increased steadily from 2001 to 2005 in Europe (shown in Figure 1 below). Therefore, financial institutions require credit scores to minimise their risk.

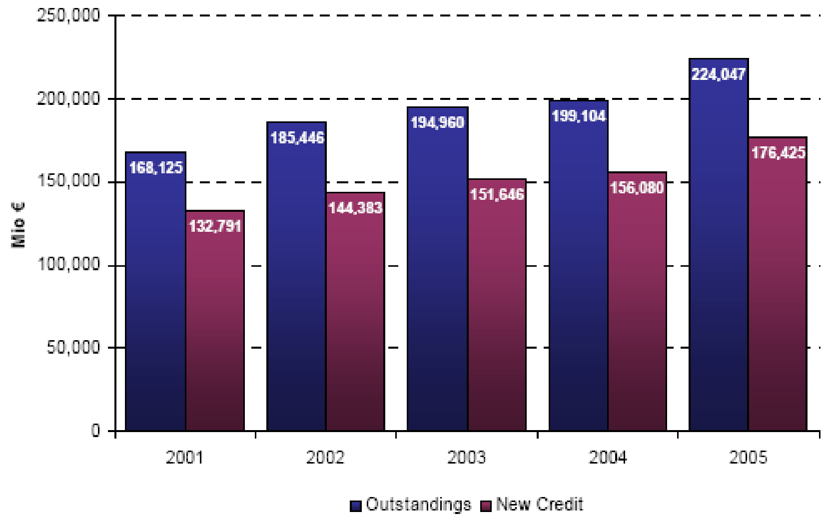


Figure 1 - Consumer credit on the rise (Eurofinas, 2005)

2.2.1. Information used by credit scoring models

Credit scores are designed to measure the risk of default by taking into account various factors in a person's financial history. The exact formulas for calculating credit scores are closely guarded secrets. FICO has disclosed that the following components and weighted contribution of each to the calculation of the credit score illustrated in Figure 2, below):

- 35% - punctuality of payment in past
- 30% - amount of debt (ratio of outstanding balances (revolving debt) to total available credit (credit limit)).
- 15% - length of credit history
- 10% - type of credit used
- 10% - recent credit and amount of credit received recently

The components above are fed as inputs to a model that produces the credit score which is ultimately a predictor of whether the consumer will meet their debt obligations or not.

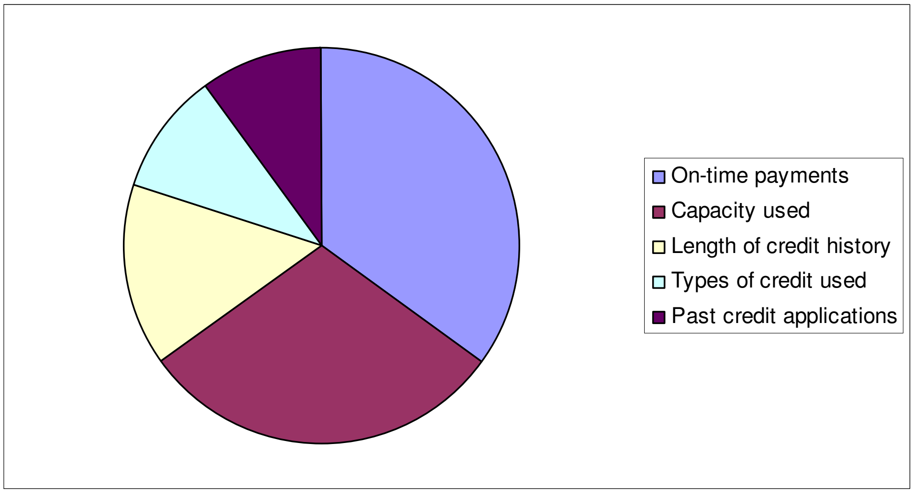


Figure 2 - Information used to calculate credit scores

2.2.2. Problems with credit scores

There are several problems associated with credit scoring. Firstly, Individuals with higher credit scores are offered different services than those with lower scores. Individuals with lower credit scores are targeted with sub-prime loans with higher interest rates. Sub-prime and predatory loans are disproportionately made to individuals with lower credit scores. Secondly, credit score use is being expanded controversially e.g. insurance companies are using credit scores to assess risk levels and loss ratios despite the lack of a causal link between a credit score and an insurance risk (Electronic Privacy Information Centre, 2003). The South African National Credit Act (NCA) of 2005 normalizes the credit market by regulating the relationship between the consumer and the credit provider. This proposed bill aims to give equal protection to all borrowers (Clayton, 2005). Thirdly, inaccuracies in credit reports can result in serious disadvantages for consumers due to the importance of credit reports.

2.3. Artificial Neural Networks

2.3.1. Introduction to ANNs

Artificial neural networks are intelligent systems based on the biological structure of the human brain and consist of many simple elements (neurons) operating in parallel.

Biological neural networks are much more complicated than the mathematical models we use for ANNs. Practical ANNs are extremely small in comparison with the human brain but their basic structure remains highly connected and distributed in nature which affords them a high degree of noise immunity and fault tolerance. ANN learning algorithms are also relatively insensitive to noise since they are aimed at identifying general trends and relationships in data and noisy training data actually helps to accomplish this in some cases.

This is particularly useful where the comprehensive models that are required for conventional computing methods are either too large or complex to accurately represent systems/signals. The credit scoring application is a good example of this. The responsibility of identifying the differences between the signal representations of different consumers is now that of the neural network.

2.3.2. Capabilities of ANNs

ANNs often employ supervised learning. For supervised learning, training data that includes both the input and the desired result (the target value) must be provided. After successful training, the input data can be presented alone to the ANN (that is, input data without the desired result), and the ANN will compute an output value that approximates the desired result. However, for training to be successful, lots of training data is required. Appropriate preprocessing of the input data to numeric values is usually required. ANNs are especially useful for classification and function approximation or mapping problems which are tolerant of some imprecision, which have lots of training data available, but to which hard and fast rules (such as those that might be used in an expert system) cannot easily be applied.

2.3.3. Level of Measurement of ANN input data

ANNs differ in the kinds of data they accept. Two major kinds of data are categorical and quantitative:

Categorical, or nominal, variables take only a finite (technically, countable) number of possible values, and there are usually several or more cases falling into each category. Categorical variables may have symbolic values (e.g., "male" and "female") that must be encoded into numbers before being input to the network.

Quantitative variables are numerical measurements of some attribute. The measurements must be made in such a way that at least some arithmetic relations among the measurements reflect analogous relations among the attributes of the objects that are measured i.e. ordinal, interval or ratio data.

The purpose of this research is classification and the various biographic, demographic and financial input data will be scaled into categorical and quantitative variables before input to the training LVQ and the trained LVQ for testing or validation.

2.3.4. The neuron model

The inputs to a neuron, shown in Figure 3, include its bias (b) and the sum of its weighted inputs (p) using inner product. Biasing is optional and not present in all neurons. The input to the neuron is then augmented by the transfer function (F) which produces a scalar output (a). Appendix A describes some of the various transfer functions that exist for neural networks. The bias is much like a weight except that it has a constrained input 1. Weights (w) and bias (b) are adjustable scalar parameters. BNN networks are described in Appendix B.

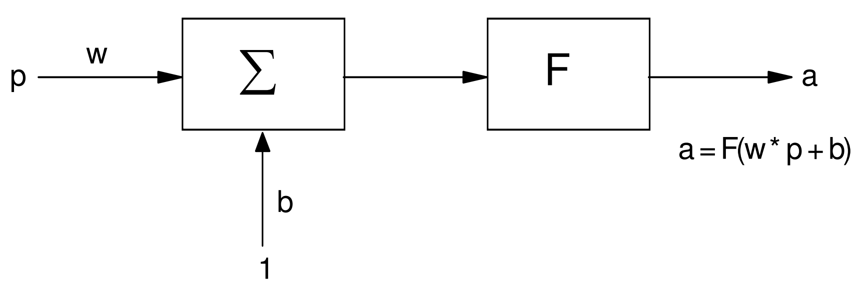


Figure 3 - Single input neuron

2.3.5. Neurons with multiple inputs

A single neuron with R inputs is shown in Figure 4 below. Individual inputs p(1), p(2),...p(R) are weighted by elements w(1,1), w(1,2),......w(1,R), and the weighted values are inputs to the summing junction. Their sum is w*p, the dot product of the row vector (w) and the column vector p. The neuron has a bias value b, which is summed with the weighted inputs to form the net input n. This sum n is the argument of the transfer function F.

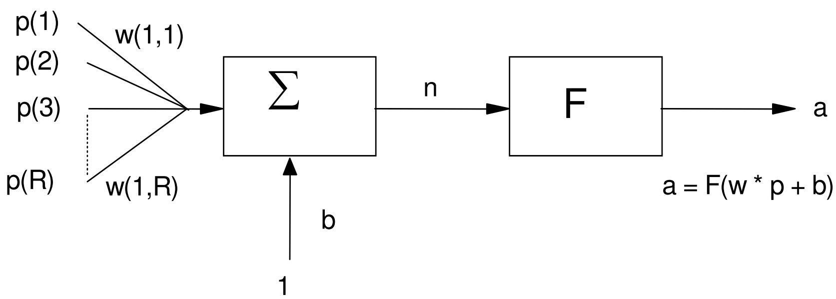


Figure 4 - Multiple input neuron

The application of credit scoring prediction with neural networks requires an input layer capable of multiple inputs per neuron.

2.3.6. ANN Training

Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. In the Figure 5 below, the network is adjusted (by adjusting the interconnections/ weights between the individual neurons), based on comparison of the target and output, until the network output matches the target. Typically, many such input/target pairs are used. The process of the ANN iterative adjustment of the weights is called ANN training. This called *supervised* network learning (Demuth and Beale, 1994) since the network is being explicitly trained with input and output target pairs. Unsupervised learning involves no target values and the network automatically categorizes the input. Figure 5, below, illustrates how supervised artificial neural networks are trained to adjust the weights of the interconnections between the neurons in order to converge to their target.

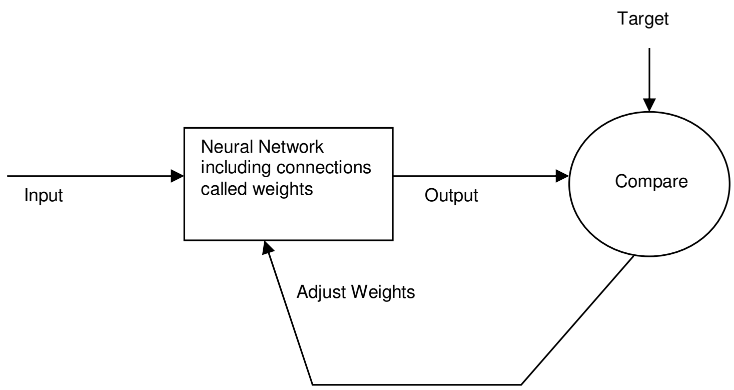


Figure 5 - Neural network training

For each input vector we calculate the output vector. The difference between the output vectors and the target vectors are the errors. The aim is to find values for the biases and weights such that the sum of the errors squared is minimized or below a specified value. ANN back-propagation training utilizes the Least Mean Square (LMS) learning rule by adjusting weights and biases according to the magnitude of errors.

Figure 6 shows how the network performance improves with training until the performance of the network matches the desired output. Function approximation has been achieved.

Batch and Incremental Learning and epochs:

Batch learning proceeds as follows: First initialize the weights. Repeatedly process **all** the training data and update the weights. Iterations of processing the training data and adjustment of the weights are known as epochs. In other words, it is the number of times the input data is fed forward through the ANN architecture during training. Epoch learning is synonymous with batch learning.

Incremental learning proceeds as follows: Initialize the weights and repeatedly process one training case and update the weights.

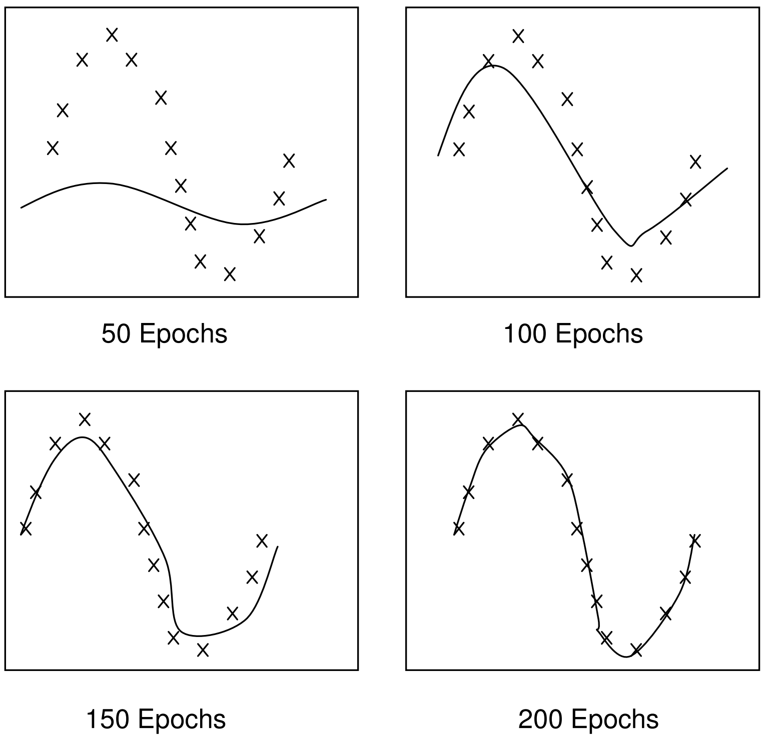


Figure 6 - Training of a back-propagation network

On-line and Off-line Learning

In off-line learning, all the data are stored and can be accessed repeatedly. Batch learning is always off-line. In on-line learning, each case is discarded after it is processed and the weights are updated. On-line training is always incremental. Incremental learning can be done either on-line or off-line (Sarle, 2002).

Once a network is trained successfully and an input vector not in the training set is presented to the network, it will tend to produce an output vector similar to output vectors associated with similar input vectors. This behaviour is called *generalization* and is what makes neural networks so powerful. This is also the testing mode of ANN operation.

2.3.7. ANN Architectures

A layer of a network is defined as follows: A layer includes the combination of the weights, the multiplication and summing operation, the bias b, and the transfer function F. The array of inputs, p will not be considered a layer. Two or more neurons may be combined in a layer, and a particular network might contain one or more such layers.

The network, shown in figure 7 below, has R inputs, with S1 neurons in the first layer. It is common for different layers to have different numbers of neurons. The outputs of each intermediate layer are the inputs to the next layer. A layer to produce the output is called the output layer. The other layers are called hidden layers. Multiple layer networks are quite powerful.

The architecture of a network consists of a description of how many layers a network has, the number of neurons in each layer, each layer's transfer function and how the layers are connected together and to the inputs. The best type of architecture to use depends on the nature of the problem to be solved.

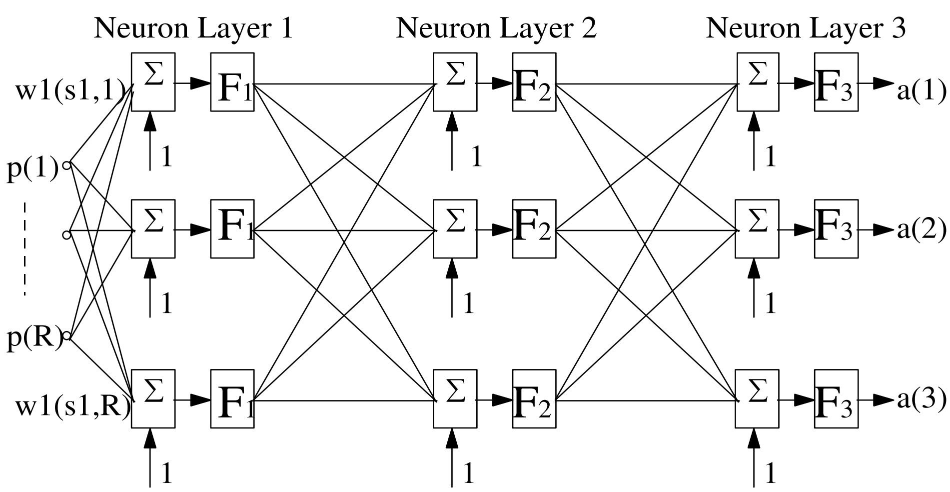


Figure 7 - Multiple layers of neurons

Aside from the number of neurons in the output layer, the number of neurons in each layer may be chosen by the designer. Linear networks are the exception where the more neurons in the hidden layer the more powerful the network.

Networks with biases can represent relationships between inputs and outputs more easily than those without biases.

2.3.8. ANN learning rate

A high learning rate leads to unstable learning which means that the network may have trouble in converging to the error minimum. If the learning rate is too small then the network will take very long to converge to the error minimum.

2.3.9. Competitive learning

The neurons of competitive networks learn to recognize groups of similar input vectors. Competitive networks may be trained with the instar learning rule where each neuron competes to respond to a specific input vector p. The neuron whose weight vector is closest to p wins and outputs a 1. All other neurons output a 0. Only the neuron that wins has its weight updated as result it is most likely to win the next time.

A limitation is that some neurons may never be allocated to an input, resulting in dead neurons. Biases are therefore used to give neurons that win rarely an advantage over those that win regularly. Each neuron is forced to classify roughly the same percentage of input vectors.

2.3.10. Self organizing maps (SOMs)

The SOM differs from competitive learning neurons in which neurons get their weights updated. Instead of updating only the winner, feature maps update the weights of the winner as well as its neighbours. Self-organising networks are good for categorisation of input vectors. SOMs utilize unsupervised learning and the user may not specify a target output category for a specific input vector.

2.3.11. Learning vector quantisation (LVQ)

LVQ is a method for training competitive networks in a supervised manner. A competitive layer automatically classifies input vectors. Classes found by the competitive layer depend only on the distance between the input vectors. If two input vectors are similar then they put into the same class. LVQ networks learn to classify vectors into target classes specified by the user.

LVQ neural networks have been used to perform the pattern recognition task of our credit scoring system. LVQs are closely related to SOMs. It is an algorithm that effectively maps similar patterns (pattern vectors close to each other in the input signal space) onto locations in the output space. LVQ is a supervised version of SOM particularly suitable for statistical pattern recognition. LVQ training requires a training set of examples of the proper network behaviour. If the input pattern is classified correctly, then move the winning weight toward the input vector according to the **Kohonen** rule. If the input pattern is classified incorrectly, then move the winning weight away from the input vector (Kohonen, 1997).

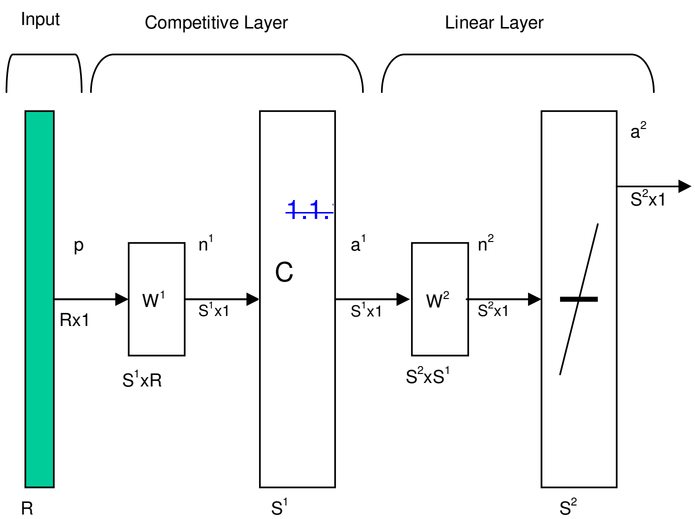
Figure 8, below, shows an LVQ network with a competitive layer followed by a linear layer. The classes learned by the competitive layer are termed subclasses and the outputs of the linear layer are termed target classes.

The output of the linear layer is given by the relationship (W^2 is the linear layer weight matrix):

$$a^2 = W^2 a^1$$

S^1 is the number of neurons in the competitive layer. This parameter is user defined. The magnitude of W^1 is dependent on S^1 according to the relationship:

$$W^1 = S^1 \times R$$



Where…

R = number of elements in input vector p

S^1 = number of competitive neurons

S^2 = number of linear neurons

Formatted: Normal

Formatted: Bullets and Numbering

Figure 8 - LVQ network architecture

(adapted from The Mathworks, 1995)

For the LVQ network, the winning neuron in the first layer indicates the **subclass** which the input vector belongs to. There may be several different neurons (subclasses) which make up each class.

The second layer of the LVQ network combines subclasses into a single class. The columns of $\mathbf{W}^2$ represent subclasses, and the rows represent classes. $\mathbf{W}^2$ has a single 1 in each column, with the other elements set to zero. The row in which the 1 occurs indicates which class the appropriate subclass belongs to e.g.

$$\mathbf{W}^2 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Subclasses 1, 3 and 4 belong to class 1.

Subclass 2 belongs to class 2.

Subclasses 5 and 6 belong to class 3.

Appendix C describes the LVQ algorithm in greater detail.

2.3.12. ANN problems

The large number of neural network applications must not obscure the fact that there are some major unsolved problems concerning neural networks. There are still no satisfactorily constructive ways to determine the optimal structure (elements as well as organization) or the learning and evaluation dynamics.

Underfitting - Training is sensitive to the number of neurons in the hidden layers. If too few neurons are present in this layer then the network will not be able to solve the problem. Thus if training a network for a long time still results in large errors, the problem is most likely a lack of hidden layer neurons.

Over-fitting – The critical issue in developing a neural network is generalization: how well will the network make predictions for cases that are not in the training set? A network that is too complex may fit the noise, not just the signal, leading to over-fitting. Over-fitting is especially dangerous because it can easily lead to predictions that are far beyond the range of the training data. The result of too many neurons is that the network appears to be trained quite easily when presented with input vectors from the training dataset but test vectors (not included in training set) do not generate reasonable results.

2.4. Summary

This chapter presented a theoretical foundation on credit scoring models and artificial neural networks.

Credit scores are used by financial institutions to evaluate the potential risk posed by lending money to consumers and to mitigate losses due to bad debt. The Fair Isaac Corporation (FICO) provides the most well-known and widely used credit score model in the United States. Credit scores computed by FICO are derived from the following factors: punctuality of payment in past, amount of outstanding debt (ratio of outstanding balances (revolving debt) to total available credit (credit limit)), length of credit history, and type of credit used and the amount of credit received recently. Problems associated with credit scoring include individuals with lower credit scores being targeted with sub-prime loans with higher interest rates, credit scoring being expanded controversially e.g. insurance companies are using credit scores to assess risk levels and loss ratios despite the lack of a causal link between a credit score and an insurance risk, and inaccuracies in credit reports that can result in serious disadvantages for consumers due to the importance of credit reports.

Artificial neural networks (ANNs) are mathematical models inspired by the biological neural networks. ANN learning algorithms are relatively insensitive to noise and are aimed at identifying general trends and relationships in data. ANNs are especially useful for classification and function approximation/mapping problems which are tolerant of some imprecision, which have lots of training data available, but to which hard and fast rules (such as those that might be used in an expert system) cannot easily be applied.

Commonly neural networks are trained, so that a particular input leads to a specific target output. The network is adjusted (by adjusting the interconnections/ weights between the individual neurons), based on comparison of the target and output, until the network output matches the target. The aim is to find values for the biases and weights such that the sum of the errors squared is minimized or below a specified value. ANN training utilizes the Least Mean Square (LMS) learning rule by adjusting weights and biases according to the magnitude of errors.

The architecture of a network consists of a description of how many layers a network has, the number of neurons in each layer, each layer’s transfer function and how the layers are connected together and to the inputs.

LVQ neural networks have been used to perform the pattern recognition task of our credit score system. LVQs are closely related to SOMs. It is an algorithm that effectively maps similar patterns (pattern vectors close to each other in the input signal space) onto locations in the output space. Learning Vector Quantization is a supervised version of SOM particularly suitable for statistical pattern recognition.

Problems associated with neural networks include the fact that there is still no satisfactorily constructive ways to determine the optimal structure (elements as well as organization) or the learning and evaluation dynamics of a neural network.
Chapter three describes the literature review conducted that relates to credit scoring utilizing ANNs.

Chapter 3

Literature Review

3.1. Introduction

Literature review forms the foundation on which research is based by providing an understanding of and insight into relevant previous research and trends that have emerged. This section describes how recent literature converges to the view that ANNs are well suited to credit risk analysis, for both consumers and company risk scoring.

The primary sources of literature used for this research are peer-reviewed academic journal publications. The secondary literature sources used in this research are books, research reports, theses, magazines and the Internet.

3.2. The need for credit risk management

Existing literature suggests that an inferior credit risk assessment tool is the primary reason of enterprise bankruptcy (Crook, Thomas & Hamilton, 1994). The main purpose of credit risk assessment is to measure the default possibility of borrowers and provide a suggestion, by conducting qualitative analysis, of credit risk. South African banks are to sign a code of conduct on how they approach customers with offers of credit cards and other types of unsecured credit such as personal loans. The code was in response to negative media reports and complaints to the banking ombudsman from customers who felt they were being harassed by banks. The code comes as household-debt levels reach record highs of more than seventy percent of disposable income following a number of years of strong credit growth due to low interest rates and rising economic growth. In December 2006, the South African Reserve Bank governor threatened to increase reserve requirements unless banks stopped reckless credit practices (Gunnion, 2007). The code was signed ahead of the full implementation of the National Credit Act (NCA) on 1 June 2007, which guards consumers against reckless lending. South African banks are also under pressure by the financial sector charter to extend loans to historically disadvantaged South Africans who had previously struggled to obtain credit. With competition and a smaller market due to compliance restrictions, banks have to make accurate decisions in order not to grant credit to customers that are unable to repay their debt. Therefore, the models used to make these decisions have to be reliably accurate. The current crisis sweeping the American housing market is also indicative of the need for credit risk management. Approximately two million people, with sub-prime or high risk mortgages, have had their homes repossessed or will default on their loans in the near future. Borrowers and lenders are both losers alike as major mortgage lenders such as New Century Financial have gone bankrupt (Sherwell, 2007).

(The Credit Research Foundation, 2006) defines credit scoring as a method of evaluating the credit worthiness of customers through the implementation of a formula or set of rules. The methodology of testing the credit worthiness of the customer base via credit scoring has evolved. Studies performed by the Credit Research Foundation (CRF) revealed that current technological change is changing what credit functions are performed and how they will be fulfilled. This sophisticated technology will have to be harnessed due to economic pressures in order to place more emphasis on the management of the revenue chain process.

3.3. Models used for credit risk analysis

An extensive amount of research has been conducted on credit risk analysis models. Traditional methods applied to credit scoring include linear discriminant analysis and logit analysis. Emerging artificial intelligence (AI) techniques such as ANNs, genetic algorithm (GA) and support vector machine (SVM) have also been applied with positive results for both consumer and company credit ratings (Lai, Yu, Wang & Zhou, 2006a).

In this study, research will be conducted to test whether Artificial Neural Networks (ANNs) can be used to accurately detect and predict consumer credit ratings. The ANN could be used as a Decision Support System (DSS) to improve the accuracy of current credit rating scoring methods despite problems of incomplete data. Data availability is usually limited due to competition, confidentiality and privacy making the statistical approaches difficult to obtain a consistently good result for credit scoring. In order to improve the performance and overcome data shortage, it is therefore imperative to implement an approach that may be able to cope with these challenges. Pattern recognition related decision problems suit artificial neural networks more than to statistical techniques primarily because of the complex nature of problems, which are not well understood mathematically and involve subjectivity. In addition, such problems have qualitative and noise data. Even if the values of every input features are not known, a trained neural network will produce a response. This study is to take advantage of the flexible mapping abilities of ANNs (Lai, Yu, Wang & Zhou, 2006b).

ANNs observes and learns from data patterns and also have the ability to work with missing data. This is particularly useful where the comprehensive models that are required for conventional computing methods are either too large or complex to accurately represent systems or signals. There is no evidence of any single variable that can be accurately used to predict consumer credit default, so a non-linear system (such as an ANN) may be a step in the right direction e.g. different combinations of consumer characteristics could possess inherent predictive capabilities of the probability that they will default. Specifically, a LVQ ANN will be used as a pattern recognition model. An overview of LVQ network rules and architecture is presented in Appendix C.

ANNs could be ideally suited to the turbulent financial risk management field due to their adaptive nature – different combination of factors may be weighed differently for different consumers. Requirements may also vary over time - the system could be periodically updated by retraining with more recent data to reflect changes in consumer behaviour and the macro-environment. The ultimate aim would be online training, where the training algorithm would be capable of processing the input piece-by-piece, without having the entire input available to train the network from the start (Wikipedia, 2007). Online training with an ANN could be achieved by storing the trained networks weight matrices and training from that point on with new data. Also, the system may be able to adapt to the different environments e.g. factors that predict credit suitability could be different between developed and developing nations (South Africa), different age groups, different gender and race groups, etc.

Studies have shown that Artificial Intelligence (AI) methods achieved better performance than traditional statistical methods (Atiya, 2001). An ANN, which is an information processing device inspired by the biological nervous system, is a subset of AI (Lorenzo, Martinez & Garcia, 2000). Kumar and Haynes (2003) found the Artificial Neural Network (ANN) model to be superior to the conventional discriminant analysis model in terms of accuracy when applied to rating debt obligations. The objective of their study was to verify whether a relationship exists between financial information of a company and the ratings on debt obligations awarded by experts. The study proposed to use Artificial Neural Networks, which have a proven ability to capture hidden relationships between the dependent variable, and a set of independent variables and compare the classification and prediction results with results of conventional discriminant analysis. They found that the ANN model could be used to increase speed and efficiency of the rating process in practical applications. In addition, they found that if better-input data was provided, it could be relied upon to provide an automatic rating to a significant extent. Their testing of the ANN network performance was limited to 10 cases (13%) whilst the remaining 66 cases (87 percent) were used as the training set.

The extensive use of neural networks as powerful tools for financial business decisions is also described by Krishnaswamy, Gilbert and Pashley (2000). It is stated that materials promoting this finance technique have grown explosively. ANNs allow for classification, clustering, and pattern recognition of data. They are applied to forecast stock and commodity prices, bond ratings, foreign exchange rates, T-bills, bonds, and inflation. ANNs are proven to be helpful in security trading systems, pricing and hedging of derivatives, forecasting the financial success and demise of firms (including small firms) and predicting merger targets. Banks have utilized them for credit scoring and fraud detection (especially for credit cards), evaluating signatures, voice recognition, and re-establishing incomplete fingerprints (Demuth and Beale, 1994). Other advantages of ANNs include the ability to handle ill-structured data and noise more successfully than expert systems. Pattern recognition related decision problems suit artificial neural networks more than statistical techniques primarily because of the complex nature of problems, which are not well understood mathematically and involve subjectivity. In addition, such problems have qualitative and noise data. Even if the values of every input feature are not known, a trained neural network will produce a response. Large data sets are required to train ANNs, although ANNs can function well in the presence of missing data. ANNs are powerful for analyzing massive unstructured and complex data sets. ANNs also have disadvantages which include excessive iterations needed to train the ANN, under or over-fitting data, and difficulty interpreting results due to the "black box" nature of ANN processing.

Huang, Chen, Hsu, Chen and Wu (2004) introduced a relatively new machine learning technique, support vector machines (SVM), in an attempt to provide a model with better explanatory power than ANNs. They used BNN as a benchmark and obtained prediction accuracy around 80% for both BNN and SVM methods for the United States and Taiwan markets. This study showed that ANNs can be applied to financial risk management with similar accuracy as SVM.

Lai, Yu, Wang and Zhou (2006a) indicated that the main reason of selecting ANN reflects their ability to provide flexible mapping between inputs and outputs. Neural networks are viewed as a “universal approximator” Usually, a three-layer back propagation neural network (BNN) with an identity transfer function in the output unit and logistic functions in the middle-layer units can approximate any continuous function arbitrarily well given a sufficient amount of middle-layer units. The motivation of their study was to propose a triple-phase neural network in order to increase performance since many experimental results have shown the generalization of individual networks is not unique. Different neural networks with different settings (e.g. different network architecture and different initial conditions) may result in different classification results. However their study indicates only a modest improvement in recognition accuracy.

Jagielska and Jaworski (1996) showed that neural network technology can be used to assist experts in credit card risk assessment. Their neural network model was able to identify bad accounts in the population of applications that were previously assessed as good accounts. They divided database information samples into training, validation and testing sets. The validation set was used to optimize the ANN network architecture selection to detect over-fitting during training. Over-fitting occurs when neural network architecture selection problems result in the network finding local solutions rather than generalized solutions. A BNN algorithm was used and the validation results revealed that a network with three hidden layers, consisting of 30 neurons, 10 neurons and 2 neurons respectively, produced the most accurate classification results. They further conclude that the information contained in the credit card applications were insufficient for accurate risk assessment. They suggested that further research, investigating whether the inclusion of extra input variables (economic and credit limit changes) can improve the accuracy of prediction.

The different approaches to credit risk analysis can be classified as individual models, such as linear discriminant analysis, logit analysis, probit analysis, linear programming, integer programming, k-nearest neighbor (KNN), ANN and SVM, and hybrid models such as neuron-fuzzy systems and fuzzy SVM. Logit analysis has been the most popular statistical technique. It uses logistic regression to predict a binary outcome such as bad debt or no bad debt. Li, Shi, Zhu and Dai (2006) also proposed a data-mining approach to classify credit cardholders' behavior. The underlying approach to data-mining is to search large volumes of data for patterns using tools such as classification and clustering. The research by Li, *et al* (2006) involved dimension reduction using principal component analysis (PCA) and ensemble method based on Multi-criteria linear programming (MCLP) classification method. PCA is a useful multivariable statistic analysis method, which is an excellent linear dimension reduction technique to transform more indexes or attributes into few new ones.

Komor´ad (2002) investigated credit scoring prediction accuracy and performance on a data set from a French bank. The credit score prediction performances of the following models were compared: popular logistic regression, multi-layer perceptron (MLP) neural network and radial basis neural networks were compared. The results obtained indicated that the methods, namely the logistic regression, multi-layer perceptron (MLP) and radial basis function (RBF) neural networks give very similar results, however the traditional logit model seems to perform marginally better. This work also presents an overview of other methods used for credit scoring:

- Probit Regression
- Semi-parametric Regression
- Linear Discriminant Analysis
- Panel Data Analysis

- Hazard Regression
- Genetic Algorithms
- Linear Programming
- Treed Logits

Komor´ad (2002) did not consider LVQ ANNs in his study.

3.4. Hybrid models

In the two listed hybrid models by Lai *et al.* (2006a), the neuro-fuzzy system performed worse than that of two single AI models, i.e., ANN and SVM. This study therefore concentrates on an individual ANN model and applies the network to the South African consumer credit sector. The features of this emerging economy, its consumer behaviour and the influence of the specific macro-environment make it unique. In a perfect world, the best approach to credit scoring is combining the models, taking advantage of the strengths and, thus, creating a hybrid approach to the overall credit scoring process.

3.5. No Free Lunch theorem

It is difficult to say that the performance of one individual model is consistently better than that of another model in all circumstances (Lai, *et al.*, 2006a). They found the performance of these individual models to be problem-dependent. Of the individual models, the neural network model performs the best, followed by single SVM and logit regression. They also found that the difference between the performance of ANN and SVM is insignificant at five percent level of significance, while the difference between logit regression and ANN is significant at ten percent level of significance. Yu**,** Lai and Wang (2006) also assessed credit risk with Least Squares Support Feature Machine (LS-SFM). The results obtained showed that the LS-SFM is an effective classification tool for credit assessment. In addition, LS-SFM can deliver the users how important each feature is. Neural networks have been proved to outperform conventional statistical model in terms of classification accuracy and the limitation of reduction. However, structure selection is always the difficult part of these methods. Unsuitable structure will lead to local minima or over-fitting.

Excessive training and exposure to examples can also lead to the neural network memorizing the training set data. The ANN is then unable to generalize when applied to test data. A conventional interpretation of the no free lunch theorem (NFL) is that “a general-purpose universal optimization strategy is theoretically impossible, and the only way one strategy can outperform another is if it is specialized to the specific problem under consideration” (Ho and Pepyne, 2002: 549). Different models (algorithms) may obtain different results on a particular problem but, over all problems, they are indistinguishable. ANNs will be used in this research in an attempt to find a solution to the specific problem of consumer credit scoring in South Africa.

Lanquillon (1999) shows the dynamic aspects of neural classification in environments that are constantly changing with time. The parameters of the neural network should be updated regularly in order to retain classification accuracy else the neural network may lose performance due to structural changes. The dynamic aspects and structural changes over time are not considered in this study as the data to be obtained is not conducive to a longitudinal study.

3.6. Company risk analysis

There is also extensive literature on the study of company credit risk analysis. (Bandyopadhyay, 2006) aimed developing an early warning signal model for predicting corporate default in emerging market economy like India. Using Multiple Discriminate Analysis (MDA), a high classification power on the estimated sample and also high predictive power in terms of its ability to detect bad firms in the holdout sample was achieved. Self Organising Maps (SOM) has been used by Länsiluoto, Eklund, Back, Vanharanta and Visa (2004) to build a neural network model using the self-organizing map technique and illustrates the Industry-specific cycles and companies’ financial performance comparison.

3.7. Tentative propositions and hypotheses

The following are tentative propositions from the researcher, based on the literature review:

Hypothesis 1 - ANNs can be trained to predict consumer credit risk in South Africa utilizing their biographic, demographic and credit history as inputs to an LVQ classification/ pattern recognition ANN.

Hypothesis 2 - ANNs can be used to improve the accuracy obtained with conventional consumer credit scoring systems.

3.8. Summary

This chapter reviewed literature relevant to credit risk analysis with artificial neural networks. The primary sources of literature used for this research are peer-reviewed academic journals publications. The secondary literature sources used in this research are books, research reports, theses, magazines and the Internet.

There is an essential need for credit risk management as existing literature suggests that an inferior credit risk assessment tool is the primary reason of enterprise bankruptcy. The implementation of the National Credit Act (NCA) on 1 June 2007 in South Africa will force banks to stop reckless credit practices.

An extensive amount of research has been conducted on credit risk analysis models. Traditional methods applied to credit scoring include linear discriminant analysis and logit analysis. Emerging artificial intelligence (AI) techniques such as ANNs, genetic algorithm (GA) and support vector machine (SVM) have also been applied with positive results for both consumer and company credit ratings. This study focuses research on whether Artificial Neural Networks (ANNs) can be used to accurately predict consumer credit ratings. The ANN could be used as a Decision Support System. Pattern recognition related decision problems suit artificial neural networks more than to statistical techniques primarily because of the complex nature of problems, which are not well understood mathematically and involve subjectivity. In addition, such problems have qualitative and noise data. Past research has shown that neural network technology can be used to assist experts in credit card risk assessment.

It is difficult to say that the performance of one individual model is consistently better than that of another model in all circumstances (No Free Lunch theorem). The performance of these individual models has been found to be problem-dependent. The tentative propositions and hypotheses of the researcher, based on the literature review are therefore:

Hypothesis 1 – ANNs can be trained to predict consumer credit risk in South Africa utilizing their biographic, demographic and credit history as inputs to an LVQ classification/ pattern recognition ANN.

Hypothesis 2 - ANNs can be used to improve the accuracy of credit risk analysis.

Chapter four follows with the research methodology utilized to meet the objectives of this research endeavour.

Chapter 4

Research Methodology

4.1. Introduction

Chapter 4 provides an exposition of the investigation provided in this research report. Firstly, the objectives and aims of the research are clearly defined. This is important since the methodology is derived from the objectives of this study. Once the research objectives were defined, it was then possible to conduct the literature review. The hypotheses are based on the results of the literature review. The research problem and sub-problems are appropriately restated in this section.

The literature review converges to the hypothesis that LVQ ANN neural networks may be suited to increase the performance of credit risk analysis systems. The proposed credit risk analysis system model is described in this chapter.

The consumer credit data used to verify the accuracy of the system is described. Data acquisition, sampling methods and techniques used for the research is also discussed in detail. In addition, the issues of reliability and validity of the research methodology are reviewed.

The choice and architecture of the LVQ ANN, the number of training cycles (epochs), the training dataset, validation dataset and testing dataset are also explained in this chapter.

The methodology of this research is aligned with the research process suggested by Cooper and Schindler (no date) as displayed in Figure 9 below.

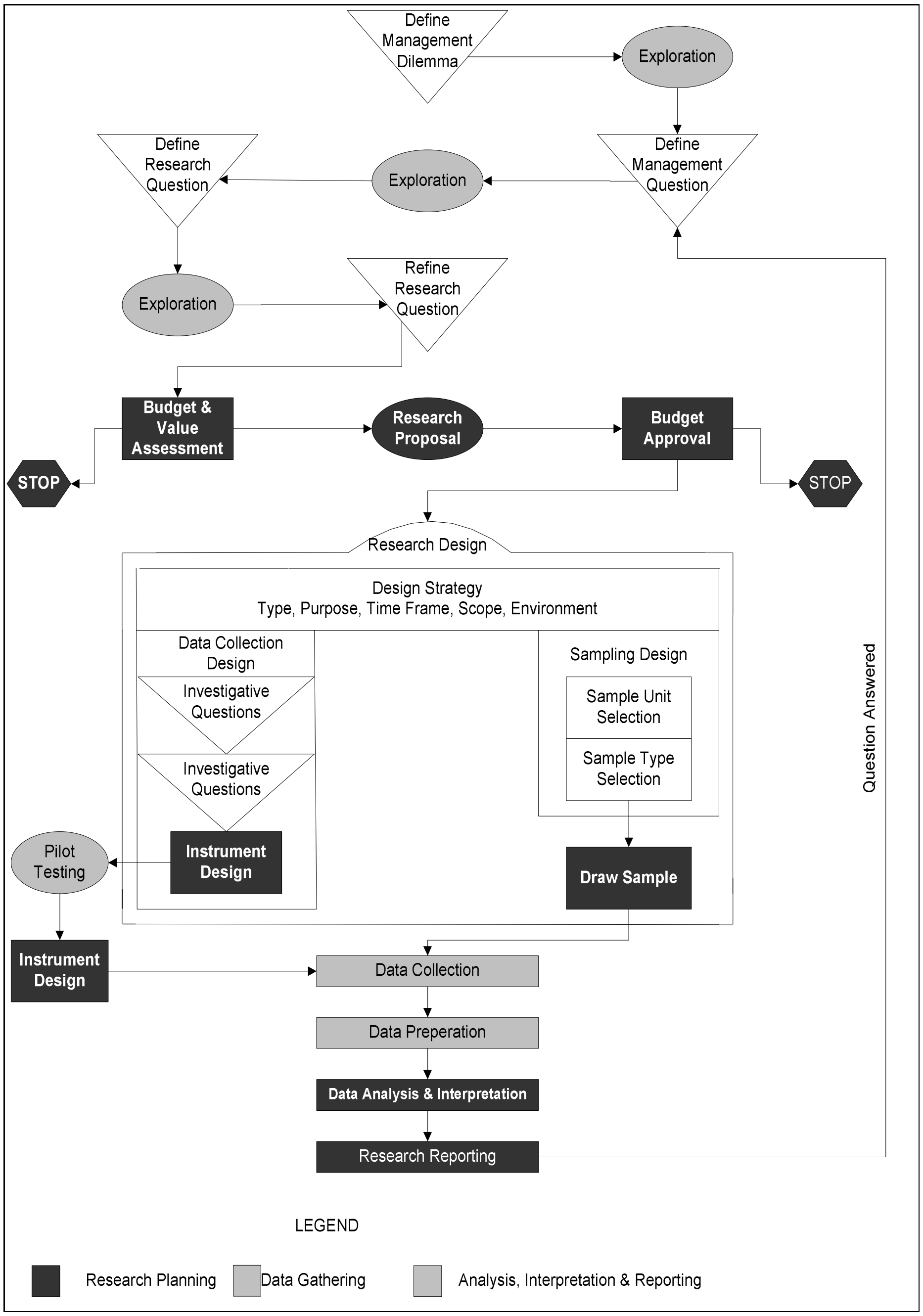


Figure 9 - The Research Process

Source: (Cooper and Schindler, no date)

4.1. Research design strategy

4.1.1. Type of research

This research work is quantitative in nature. Leedy and Ormrod (2005) describe quantitative (also called traditional or experimental research) research as research used to answer questions about relationships among measured variables with the purpose of explaining, predicting and controlling phenomena. In contrast, qualitative research is typically used to answer questions about the complex nature of phenomena, often describing and understanding phenomena from the participant's point of view.

4.1.2. Timeframe of research

The total timeframe of the research is the one calendar year of 2007. This includes the research proposal submission, literature review, data collection, data analysis and the final reporting of results and conclusions. The data used for analysis consists of all consumer and SME consolidated accounts that existed at 30 April 2007.

4.1.3. Environment and scope of research

The research data was obtained from a leading South African banking institution and is limited to consumer and SME credit analysis. The sample frame consists of the data of all the bank's current customers as at 30 April 2007. These customers have been rated/ classified into risk categories by the bank's Behavioural Scoring System (BSS).

The bank uses a complex method for behavioural scoring that includes criteria such as the amount of time that the individual has been a customer of the bank, etc. A behavioural score predicts the likelihood of an account going 'bad', based on payment history, usage, delinquency and timing characteristics i.e. risk scores are designed to predict the future performance or behaviour of an account. Therefore in determining a prediction of future behaviour for the customer and all his/her accounts, the bank's financial experts look at behaviour over a six month period i.e. the principle "the future is like the past" is applied. Subjectivity plays no part in the derivation of these scores. The scores are derived objectively utilising various statistical models that always treat all like accounts/clients **consistently** the same and does not cater for exceptions. A customer's behavioural score, which is a function of the behaviour of the account holder, is **systemically irreversible**. A behavioural score is calculated on a monthly basis.

The bank's BSS uses scoring at two different hierarchical levels: the account level score and the customer level score. At the account level, Performance Indicators (PI) which range from 0 to 9 are used to indicate the account risk. A higher rating indicates a higher the risk. A PI rating of 0 indicates accounts that were not scored e.g. closed, deceased, new accounts (open for less than 6 months), etc. Examples of some typical characteristics that may be used in account level behavioural scores are illustrated below:

Days in excess for the last six months (weighting 12)
Overdraft Limit Utilisation as a percentage of Minimum Balance Last 6 Months (weighting 18)
Days in Debit Last 6 Months (weighting 11)
Current Minimum balance (weighting 8)
Current Days in Excess (weighting 11)
Current Days in Credit (weighting 6)
Average Credit Turnover Last 6 Months (weighting 11)
Months Account Open (weighting 13)
Payment Reversals Last 6 months (weighting 10)

At the customer level, Customer Level Risk Grades (CRGs) are used. CRGs range from 0 to 5 and reflect the customer's current overall risk where the higher the rating the higher the risk i.e. a CRG rating of 1 is very low risk and a CRG rating of 5 is very high risk. A CRG of 0 indicates that a customer has not been scored e.g. closed or deceased accounts, etc. A CRG 5 implies very high risk (CRG rating of 4 defaults to a CRG rating of 5 if the worst customer account PI equals to 9).

The account and customer level scores are hierarchical. All account PI's are rolled up to Customer level where the CRG is determined using customer scorecards for consumers and SME's. These customer scorecards use customer level attributes **together** with account level & PI attributes where these attributes carry various weightings on a scorecard. Therefore, numerous attributes, both on account and customer level influence the CRG. Examples of some typical characteristics that may be used in customer level behavioural scores are:

Worst current account PI (weighting 30)
Worst revolving credit account PI (weighting 18)
Worst fixed term account PI (weighting 19)
Time with bank (weighting 6)
Percentage Un-secured lending to exposure (weighting 11)
Customer age (weighting 6)

This research focuses on using ANNs for credit scoring at the customer level. Customers with high credit risk can be automatically detected by the model at an early stage. These customers, suspected of delinquency by the ANN, can now be prioritized for further analysis by the banks financial experts.

4.2. Restatement of the research problem

The research design strategy originates from the research problem. The research problem was previously stated in section 1.3. The main research problem is to test whether LVQ artificial neural networks can be applied accurately to South African consumer credit risk analysis.

Sub-problems include the following:

a) Are there relationships between customer demographic and biographic characteristics and their credit risk grades?
b) Can LVQ ANNs successfully determine consumer credit rating classifications from the limited biographic, demographic and historic credit information available from banking institution databases?
c) How does the architecture and number of training cycles of the LVQ network affect the accuracy of the classification system?

d) How does the accuracy of the proposed LVQ network compare to existing classification systems currently used by the financial institutions?
e) How does the accuracy of the proposed LVQ network compare to actual consumer repayment behaviour? The actual repayment behaviour will be compared to the prediction of the ANN.

The hypotheses of this research study are based on the research problem and the literature review on the use of ANNs as credit score models. In order to gain better understanding of the methodology implemented, it is important to first explain the LVQ/ ANN system model before explicitly stating the hypotheses to be tested.

4.3. The LVQ/ANN system model

The LVQ system block diagram is shown in Figure 10 below in order to explain the rationale of hypotheses selection for this research study.

4.3.1. LVQ training mode/ phase

The consumer data (behavioural (e.g. past credit history), biographic and demographic (e.g. gender and marital status) are pre-processed and fed as inputs to the classifier (LVQ ANN). These individual characteristics, e.g. gender and marital status, are fed as elements of a single vector for each consumer i.e. consumer 1 will have a corresponding input vector 1 which is comprised of nominal, ordinal and ratio variables such as marital status, age and income per annum respectively. The output of the system is the risk category of the consumer. The output credit risk category will be specified to the LVQ system, for each consumer input vector, during the training phase. It is during this phase that the ANN continually adjusts the weights in order to establish relationships between consumer input attributes and their credit scores. The total number of epochs (iterations) is user-determined. The ANN is 'learning' through experience. Once training has been completed the ANN stores the final adjusted weight and bias values in a weight matrix.

The data used during the training mode is known as the training dataset. The training dataset is typically seventy percent of the total sample dataset. Lai, *et al*.

(2006a) and Komo'rad (2002) used approximately seventy percent of the sample as the training data set. The remaining thirty percent of the sample was used for validation and testing the ANN. The success of the ANN training mode is determined with data from the training dataset. The output of the ANN is then correlated with the actual credit score outputs from the training dataset.

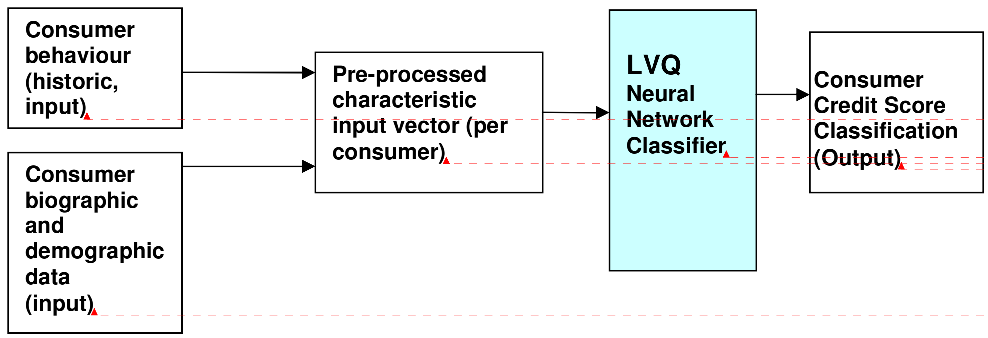


Figure 10 - Schematic of artificial neural network model for credit risk prediction

Training will be deemed successful once the ANN can successfully categorize the training data. The number of training epochs will be increased and the structure of the ANN will be amended to include more layers or neurons per layer if required.

4.3.2. LVQ validation mode/ phase

The phase that follows LVQ training is called validation. ANNs can suffer from over-fitting and under-fitting due to non-optimal selection of the ANN network architecture. This will be minimized with the use of a validation data set. The LVQ network architecture will consist of several layers: An input layer, hidden layers and an output layer. In addition, each layer can have a various number of neurons (elements). Although there are a few rule-of-thumb recommendations, there is no precise method for optimal selection. The architecture will be varied and optimized with the use of the validation data before the final architecture to be used on the test dataset is confirmed.

The validation mode requires a validation dataset. The validation dataset is typically 10 percent of the total sample dataset. This is a set of data that is used to optimize the ANN structure to fit the specific problem (credit scoring in our case). The number of iterations during training can also be optimized with the validation dataset. This data is not presented to the ANN during the training mode. The output is also not specified and the ANN has to use the LVQ testing algorithm to determine the output. Re-training (returning to the training mode) will be necessary each time the architecture is changed.

4.3.3. LVQ test mode/ phase

This is the final phase of testing the LVQ credit scoring performance. The ANN has been validated and the optimal structure has been determined. The dataset used is different from both the training and validation dataset. This is typically 20 percent of the total sample data set.

Ripley (1996: 354) has provided the following definitions:

Training set:

> A set of examples used for learning, that is to fit the parameters [i.e., weights] of the classifier.

Validation set:

> A set of examples used to tune the parameters [i.e., architecture, not weights] of a classifier, for example to choose the number of hidden units in a neural network.

Test set:

> A set of examples used only to assess the performance [generalization] of a fully-specified classifier.

4.4. Research data analysis methodology

This section describes the acquisition of the research data, description of the dataset and the sampling methods utilized to select the data. It is important that the data is described before the hypotheses are presented.

4.4.1. Data Acquisition

Consumer credit data has been obtained from one of South Africa’s largest banking institutions. This data is used as the target population and a sampling frame for this study. This data will be used to derive the sample set and to test the hypotheses.

4.4.2. Datasets requested

The following data sets were requested (per consumer) from the banking institution to train and test the accuracy of the LVQ ANN system:

a) Inputs to the system: Consumer demographic and biographic (basically as much information as is available that will allow the ANN to find an association between the consumer characteristics and their credit rating. This data will be used anonymously (no names, addresses, ID numbers, telephone numbers, etc are to be disclosed). Random symbols will be used to code each subject. The composition of input vector is determined by demographic, biographic and behavioural criteria e.g. age and sex are biographic fields whilst previous credit history and number of defaults are behavioural attributes that could be predictive of the customer’s credit risk. Yobas, Crook and Ross (2000) used the following characteristics as variables that composed the input vector:

Applicant's employment status
Years at bank
Home mortgage value
Number of children
Years at present employment
Residential status
Types of account
Other cards held
Outgoings
Estimated value of home
Home phone
Applicant's income
Spouse's income
Major credit cards held

Jagielska and Jaworski (1996) included age, postal code, time at address, home phone number, residential status, occupation, time currently employed, time previously employed, number of loans, identification of the applicant, additional cardholder, bureau status, gender, number of dependants, marital status, spouse employed, referee phone, time bank account open, net assets, income, bank balance, surplus and the number of defaults in their study.

In addition to the characteristic variables listed above, the following data was also requested from the financial institution in order to improve the predictive qualities of the neural network model i.e. its ability to generalize:

Age
Sex
Nationality
Marital status
Spouse employment status
Number of dependents
Number of siblings
Preferred language of correspondence
Home language
Population group
Province of residence
Time at residence
Occupation, Job title & date of assumption of current post
Income per annum
Economic sector
Educational qualifications
Previous credit history
Account balance and surplus
Number of defaults
Outstanding debt
Property Owner
Motor Vehicle Owner

Ideally, data on all the fields listed above should be used as components of the input vector. This is not possible as the bank has not recorded all the fields of data for each customer. Further, there is much missing information from certain fields of customers in the data provided.

b) The following is needed in order to train the LVQ classifications (outputs of the system = credit rating assessments)

The credit rating that was applied by the bank prior to the credit approval.
Actual consumer settlement performance after the rating was assigned.
This data will be formatted into an output vector per consumer.

It has also been requested that the institutions supply the method used to calculate the consumer credit scores that they applied.

4.4.3. Data/ sample frame obtained

Data was granted by a leading South African bank for purpose of this research. The data is extracted from the bank's various databases and is presented at the customer-level (including all account level data). Table 1 shows the characteristic input variables available for this research analysis are a combination of behavioural, demographic and biographic types:

Table 1 - Sample frame data fields

Date of Birth	could indicate which generation the customer belongs to e.g. baby boomer, generation X, etc
Nationality	Different country citizens may present different risks (economic effects, etc)
Occupation	May indicate ability to repay debt
Age of Business	Will not be used as this research focuses on consumer credit risk
Race	May reflect the effects of previous social injustices or current policies e.g. affirmative action, black economic empowerment, etc

Date of Assumption of account	Will not be used, “time with bank” will be used instead, this field has much missing data
VIP Indicator	prominent member of society
Staff Indicator	employee of the financial institute
Employment Type	Professional, retired, housewife, etc
Residential Status	indicates if the customer is an owner of property
Customer Age (years)	Different age groups could exhibit different risk behaviour. could indicate which generation the customer belongs to e.g. baby boomer, generation X, etc
Marital Status	is the customer single, married in community of property, etc
Gender	male or female
Preferred Language of Communication	limited to English or Afrikaans
Consumer or Business customer	Will not be used in this research as it is limited to consumers only
Time With Bank (in months)	An indicator of stability
Annual Income	Capacity to repay
Total Credit Turnover (for the last three months)	A behavioural indicator
Total Debit Turnover (for the last three months)	A behavioural indicator
Ceiling on Short Term Loan	A limit that is set , by the lending institute, using the customers profile e.g. salary/ annual income, assets, etc.
Total Outstanding Balance	This is the current outstanding amount of present customers but it could also be the amount of credit being applied for by new customers (to assess new applicant risk)

The corresponding output classification, for each customer, is the banks software system risk grading:

Risk Grade (CRG rating between 1 and 5)

4.4.4. Sampling

The sampling frame to be used is the institution's total active customer accounts as at 30 April 2007. The data will be sampled in order to obtain a dataset that is representative of the population. This is necessary in order to reduce the computational cost and processing time of the neural network. Sampling is also necessary to eliminate the adverse effects of over-fitting the neural network.

This dataset consists of 2680159 customer accounts. 2350016 of these are consumer accounts and the remaining are SME accounts. This research is limited to the consumer accounts. This data is analysed using Microsoft Access as it contains too many records to be viewed by Microsoft Excel (truncates data after the 65536th customer record). This further motivates the need for sampling.

Identifying a sufficient **sample size** is important. Larger sample sizes are desirable as they are more representative of the population. Sample size is inversely proportional to sampling error and a large sample is desirable. The Bayesian approach for sample size determination may also be applied in order to select a sample size that maximizes the difference between the expected payoff of the sample information and the estimated cost of sampling (sampling error). Leedy and Ormrod (2005) have provided the guidelines in Table 2 for practical sample size selection:

Table 2 - Sample size selection

Population Size	Sample Size
Small population < 100	**Use entire population**
Approximately 500	**250 (50%)**
Approximately 1500	**300 (20%)**
Large populations > 5000	**400**

A probabilistic sampling method will be used since a statistical evaluation of the sampling error can be undertaken (Diamantopoulos and Schlegelmilch, 2005). Popular probabilistic sampling methods include simple random sampling, stratified sampling and systematic sampling. A combination of stratified and systematic sampling will be used since the data is a complete listing of the entire credit card account population.

Stratified sampling will be used to choose members of the population randomly from different segments (strata) of the overall population. Each stratum could be sampled in proportion to its size in the overall population (proportionate stratified sampling) to prevent sample members of different strata from having disproportionate chances of being selected (ensures representation of the minority classes). Whilst this method improves the probability of each strata being represented, it may still lead to some strata having very few members in the training sample. Therefore, disproportionate stratified sampling is used for this research and samples will be selected according to credit score strata to give a greater balance in the sample amongst risk categories. This is done to prevent a sample that contains either a disproportionate amount of high risk or low risk customers. Yu, *et al*. (2006) showed that imbalanced distribution of training samples results in the poor generalization performance of the minority class. Thus, choosing an appropriate sampling method is critical in imbalanced classification problems. Figure 11 shows the population strata distribution.

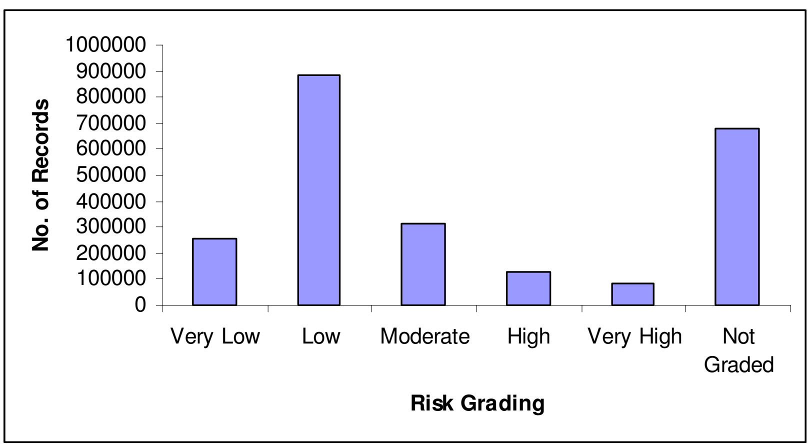


Figure 11 - Risk grading of sample frame

The sample frame has 258420 records labelled “Very Low Risk”, 884246 records labelled “Low Risk”, 314657 records labelled “Moderate Risk”, 130816 records labelled “High Risk” and 82005 records labelled “Very High Risk”. This is a total of 1670141 consumer records that have been risk graded. 679872 have not been risk graded. A sample of 400 units will be obtained from each stratum (risk grades) which will result in a total sample size of 2000. This satisfies the criteria for sample size selection (based on population size) described in Table 2. It also conforms to the normal sample size used for credit score card development as specified by Jagielska and Jaworski (1996).

Systematic sampling will be used, on each stratum, to obtain the sample dataset once the population has been stratified according to the criteria. The total sample size (n) will be predetermined and limited to 2000. The sampling interval will be determined by N/n, where N is the total population size of the credit card accounts (per stratum). Figure 12 below illustrates how the training, validation and test datasets will be obtained.

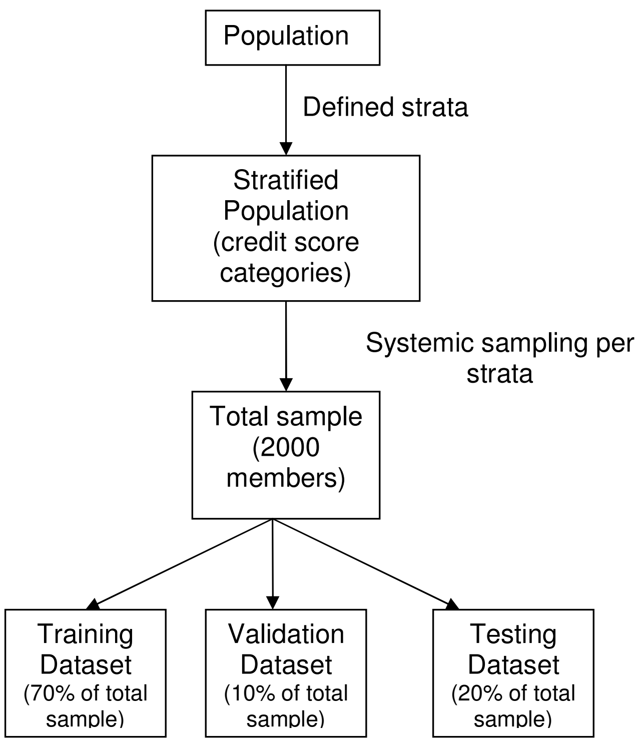


Figure 12 - Sampling method utilized

The procedure used to obtain the systemic sampling parameters for the sample of the “Very Low Risk” category is shown below:

N = total population size = 1670141 consumer records
N_{VLR} = population size of the “Very Low Risk” graded customers = 258420
n_{VLR} = required sample size of the “Very Low Risk” graded customers = 400
Therefore sampling interval of the “Very Low Risk” strata is:

$$k = N_{VLR} / n_{VLR} = 258420/400 = 646$$

A random number is generated between 1 and 646 and this is the 1st sample record. Then take every kth unit from this random number (Trochim, 2006). The same method is used to obtain samples from the remaining strata (shown in Table 3).

Table 3 - Systemic sampling intervals

Strata	Population Size per strata (N)	Required sample size (n)	Sample interval per strata (k=N/n)	Random starting record (bet 1 & k)
Very Low Risk	258420	400	646	327
Low Risk	884246	400	2210	1883
Moderate Risk	314657	400	786	691
High Risk	130816	400	327	239
Very High Risk	82005	400	205	75

The random starting record was generated in Microsoft Excel, using the RANDBETWEEN function.

4.4.5. Data preparation for analysis

An inspection of the sampled data reveals the need for data cleaning. The objective of data cleaning is to ‘identify omissions, ambiguities, and errors in the responses’ (Diamantopoulos and Schlegelmilch, 2005). A key characteristic input required for training of the ANN to recognise credit risk classes is the customer’s annual income. An inspection reveals that some data in this field is inconsistent: it seems that where data is missing in this field, it has been replaced by the value zero. This causes ambiguity since missing data from the annual income field cannot be differentiated from genuine cases where the customer's income is presently zero. The population frame has thus been limited to those cases where the annual income is non-zero. This has necessitated the need to re-calculate the systemic sample interval in order to maintain the original sample size (shown in Table 4).

Table 4 - Systemic sampling intervals revisited

Strata	Population Size per strata (N)	Required sample size (n)	Sample interval per strata (k=N/n)	Random starting record (bet 1 & k)
Very Low Risk	255941	400	639	327
Low Risk	158906	400	397	267
Moderate Risk	28670	400	71	19
High Risk	68779	400	171	65
Very High Risk	8217	400	20	1

Data cleaning is also requiring for a further field: customers time with bank. An arbitrary value 99999 is used where data was not available for this field.

4.4.6. Data coding/ scaling for input purposes

Data coding is required to transform it into a computer-readable format. This coding to a numerical scale is necessary for ANN analysis and descriptive statistical purposes e.g. Gender is coded 1 to indicate males and 2 to indicate females. The code book used is shown in the Table 5 below.

Table 5 - Data code book

Variable name	Value label	Missing value	Level of measurement
Gender	Male = 1 Female = 2	-	Nominal
Age category	0 – 20 = 1 21 – 30 = 2 31 – 40 = 3 41 – 50 = 4 51 – 60 = 5 61 – 70 = 6 71 – 80 = 7 81 – 90 = 8	-	Ordinal
Race group	White = 1 Black = 2 Coloured = 3 Asian = 4	5	Nominal
Marital status	Married in COP = 1 Married out of COP = 2 Single = 3 Divorced = 4 Widowed = 5 Tribal Marriage = 6	7	Nominal
Employment status	Professional = 1 Self Employed = 2 Retired = 3 Employed (No type/ unknown) = 4 Housewife = 5 Unemployed = 6	-	Nominal
Residential status	Owner of property = 1 Not known = 2	-	Nominal
Time with the bank (years)	0 – 1 = 1 1 – 2 = 2 2- 3 = 3 3 – 4 = 4 4 – 5 = 5 5 – 6 = 6 6 – 7 = 7 7 – 8 = 8 > 8 = 9		Ordinal

Variable name	Value label	Missing value	Level of measurement
Annual income category (x R1000)	0 – 20 = 1 21 – 40 = 2 41 – 60 = 3 61 – 80 = 4 81 – 100 = 5 101 – 150 = 6 151 – 200 = 7 201 – 250 = 8 251 – 500 = 9 501 - 1000 =10 1001 – 1250 = 11 1251 – 1500 = 12 1501 – 2000 = 13 > 2000 = 14		Ordinal
Nationality	South African (ZA) = 1 Not Applicable (NA) = 2	3	Nominal

4.4.7. Data analysis techniques

The two main kinds of ANN learning algorithms are supervised and unsupervised. In supervised learning, the correct results (target values, desired outputs) are known and are given to the ANN during training so that the ANN can adjust its weights to try matching its outputs to the target values. After training, the ANN is tested by giving it only input values, not target values, and seeing how close it comes to outputting the correct target values. In unsupervised learning, the ANN is not provided with the correct results during training. Unsupervised ANNs usually perform some kind of data compression, such as dimensionality reduction or clustering.

This research project will utilize a supervised learning algorithm. Learning Vector Quantization (LVQ) is a supervised learning algorithm and will be used to try to partition the space represented by a group of vectors into a predefined number of prototypes i.e. the consumer characteristics will be mapped to a corresponding credit classification rating.

The phases that are involved in solving this type of problem are the training phase and the classification/testing phase. During the training phase the input vector along with the known classification are cycled through the network repeatedly and in this process the network learns the mapping relationship between the input vectors attributes and the class associated with it (the credit rating). This relationship is stored in the form of weights. During the classification phase, it uses these weights to arrive at the classification of an unknown input vector.

LVQ is a method for training competitive networks in a supervised manner. In supervised learning, the correct results (target values, desired outputs) are known and are given to the ANN during training so that the ANN can adjust its weights to try matching its outputs to the target values. The data presented to the network in this stage is called the training data. After training, the ANN is tested by giving it only input values, not target values, and seeing how close it comes to outputting the correct target values. Supervised neural networks are required to learn an input-output mapping from existing data. A competitive layer automatically classifies input vectors. Classes that the competitive layer finds depend only on the distance between the input vectors. If two input vectors are similar then they are put into the same class. LVQ networks learn to classify vectors into target classes specified by the user.

LVQ neural networks will be used to perform the pattern recognition task of the consumer credit classification system. LVQs are closely related to Self-Organizing Maps (SOM). It is an algorithm that effectively maps similar patterns (pattern vectors close to each other in the input signal space) onto locations in the output space (Kohonen, 1997). Learning Vector Quantization is a supervised version of SOM particularly suitable for statistical pattern recognition. Appendix C contains an overview of the LVQ rules to be used for analysis and illustrates the LVQ network architecture.

Several software packages exist for ANN analysis. The MATLAB neural network toolbox and simulation environment, by Optimum Solutions, will be used for analysis of the data in this study. Another example of commonly available software for neural network simulation is the Neuralyst function available on Microsoft Excel.

4.4.8. Data pre-processing and level of measurement

The input range of the data will be scaled. ANNs differ in the kinds of data they accept. Two major kinds of data are categorical and quantitative:

Categorical variables take only a finite (technically, countable) number of possible values, and there are usually several or more cases falling into each category. Categorical variables may have symbolic values (e.g., "male" and "female") that must be encoded into numbers before being input to the network e.g. males will be categorized as 1 and females 0.

Quantitative variables are numerical measurements of some attribute. The measurements must be made in such a way that at least some arithmetic relations among the measurements reflect analogous relations among the attributes of the objects that are measured.

The purpose of this research is classification and the various biographic, demographic and financial input data will either be scaled into categories or retained as quantitative variables before input to the training LVQ and the trained LVQ for testing or validation.

4.4.9. Data reliability and validity

The data to be studied is envisaged to contain only those cases that have been granted credit. The limitation of this approach is that there may have been consumers that may have been denied credit incorrectly. There is no means to test this aspect since the data is not available and it is not possible to monitor those that have not been granted credit. However, as a rule of thumb in most cases, the typical credit scoring process rejects 10 percent of the account base, approves 75 percent of the account base and refers the other 15 percent to an analyst for further review (The Credit Research Foundation, 2006). Therefore, only a small percentage of applicants were rejected.

4.4.10. Statistical tests and presentation of results

Firstly, descriptive statistics will be used to describe the sample data. This will include absolute frequency, percentage distribution and summary statistics of chosen key customer characteristics. Secondly, appropriate statistical tests will be used to determine if relationships exist between key demographic biographic customers characteristics and their credit risk categories. The statistical tests that are used are the **Mann-Whitney *U* test** and the **Kruskal-Wallis one way ANOVA** tests. The strength and direction of association between these characteristics will be determined if relationships are found to be significant. **Cramer's *V*** and **Spearman's rank order correlation** (rho) will be used (where the level of measurement of both variables is at least ordinal).

In addition to statistics that describe the data and their association to the credit risk grade obtained from the financial institution, the accuracy of credit risk assessment of the LVQ network will be determined. According to Lai, *et al.* (2006a), Komor´ad (2002) and Jagielska and Jaworski (1996) there are three typical criteria that may be used to evaluate the performance accuracy of credit risk assessment classification:

$$Type\ I\ accuracy = \frac{number\ classified\ as\ bad\ credit\ risks\ by\ neural\ network}{actual\ bad\ credit\ risks} \tag{1}$$

$$Type\ II\ \ accuracy = \frac{number\ classified\ as\ good\ credit\ risks\ by\ neural\ network}{actual\ good\ credit\ risks} \tag{2}$$

$$Total\ \ accuracy = \frac{number\ classified\ correctly\ by\ neural\ network}{number\ of\ \ evaluation\ sample} \tag{3}$$

The accuracy obtained by the LVQ ANN will be compared to that achieved by similar studies where ANNs have been used for credit scoring.

4.5. Hypotheses

"A hypothesis is a logical supposition, a reasonable guess, an educated conjecture. It provides a tentative explanation for a phenomenon under investigation. It may direct your thinking to possible sources of information that will aid in resolving one or more sub-problems and, in the process, the principal of the research problem" (Leedy and Ormrod, 2005: 4).

Null hypothesis involving relationships between 2 variables X and Y:
$H_{o:}$ no relationship (i.e. zero association) between X and Y

Alternative Hypothesis:
$H_{1:}$ There is a relationship between the 2 variables (non-zero association)

The first four hypotheses relate to sub-problem a) Are there relationships between customer demographic and biographic characteristics and their credit risk grades?

4.5.1. Hypothesis 1

Hypothesis 1: Is there a relationship between the gender of a customer and their credit risk grade?

Null hypothesis:
$H_{o:}$ no relationship (i.e. zero association) between the gender of a customer and the actual credit score.

Alternative Hypothesis:
$H_{1:}$ There is a relationship between the gender of a customer and the actual credit score.

Statistical test to be used: Mann-Whitney U test since gender has only two groups (male and female) with nominal variables. Cramer's *V* can be used to test for the strength of association if a relationship does exist.

4.5.2. Hypothesis 2

Hypothesis 2: Is there a relationship between the age group of a customer and their credit risk grade?

Null hypothesis:
$H_{o:}$ no relationship (i.e. zero association) between the age group of a customer and the actual credit score.

Alternative Hypothesis:
$H_{1:}$ There is a relationship between the age group of a customer and the actual credit score.

Statistical test to be used: Kruskal-Wallis one-way ANOVA test since the age category has several (more than 2) groups. Spearman's rank order correlation can be used to test for the strength of association if a relationship does exist since both age group and credit risk grade have ordinal level of measurement.

4.5.3. Hypothesis 3

Hypothesis 3: Is there a relationship between the marital status of a customer and their credit risk grade?

Null hypothesis:
$H_{o:}$ no relationship (i.e. zero association) between the marital status of a customer and the actual credit score.

Alternative Hypothesis:
$H_{1:}$ There is a relationship between the marital status of a customer and the actual credit score.

Statistical test to be used: Kruskal-Wallis one-way ANOVA test since the marital status has several (more than 2) groups. Cramer's *V* can be used to test for the strength of association if a relationship does exist.

4.5.4. Hypothesis 4

Hypothesis 4: Is there a relationship between the race group of a customer and their credit risk grade?

Null hypothesis:

$H_{o:}$ no relationship (i.e. zero association) between the race group of a customer and the actual credit score.

Alternative Hypothesis:

$H_{1:}$ There is a relationship between the race group of a customer and the actual credit score.

Statistical test to be used: Kruskal-Wallis one-way ANOVA test since the race group has several (more than 2) groups. Cramer's *V* can be used to test for the strength of association if a relationship does exist.

Similarly, relationships between other characteristic variables and credit risk scores could be investigated.

The next set of hypotheses tests the suitability and accuracy of the LVQ ANN as a credit scoring model.

4.5.5. Hypothesis 5

Hypotheses 5 will be rejected or accepted based on whether the LVQ network architecture affects the accuracy of predicting the credit score. The number of hidden layer neurons is varied during this test.

Null hypothesis:

$H_{o:}$ The architecture of the LVQ has no bearing on the accuracy of the credit score model.

Alternative Hypothesis:

$H_{1:}$ There is a relationship between the architecture of the LVQ and the accuracy of the credit score model.

4.5.6. Hypothesis 6

This hypothesis tests the accuracy of the LVQ ANN on predicting the credit score category on the test dataset. The accuracy of the system is bench-marked against previous work in the field and the systems accuracy as described in the equations of section 4.4.10.

Null hypothesis:

$H_{o:}$ no relationship (i.e. zero association) between the actual credit score and the output of the LVQ ANN (of the test dataset).

Alternative Hypothesis:

$H_{1:}$ There is a relationship between the actual credit score and the output of the LVQ ANN (of the test dataset).

4.6. Limitations of this research report

This research is dependant on credit scoring information received from financial institutions. Data of this nature is very confidential and institutions are very sceptical to divulge records (even if coded to maintain anonymity).

The study also depends on accuracy of the credit scores assigned by the institutions. Credit scoring methods aren't easily obtainable due to the competitive nature of the credit bureau industry. (Clayton, 2005) describes the large amount of inaccuracies in credit reports.

Credit scores are often obtained by financial institutions from 3rd party credit bureaus or consultants. The actual information used to calculate the credit score is normally not available with the current account performance. This presents a problem with comparing the ANN performance with that employed by the financial institution. This work therefore assumes that ANNs can be used as an additional support system, in conjunction with the existing system, to improve the ability to predict credit default.

The ANN system is based on training with the use of past data on the premise that the past predicts the future. Changes in attitudes, economic conditions, etc. over time will affect the accuracy of the system.

The US national credit act does not permit discrimination of creditors based on demographic factors. This research aims to show whether these factors can be predictors of default, without possible practical implementation.

4.7. Summary

This chapter provided an exposition of the investigation in this research report. The methodology of this report is based on the objectives of the study. Once the research objectives were defined, it was then possible to conduct a literature review.

The research design strategy described the type of research as quantitative in nature, spanning the 2007 calendar year. The research data is obtained from financial institutions in South Africa and is limited to consumer credit analysis.

The LVQ system block diagram and the modes of operation are used to explain the rationale of hypotheses selection for this research study.

The acquisition of the research data, description of the dataset and the sampling methods (combination of stratified and systematic) to select the data are described before the hypotheses are presented. Data pre-processing, reliability and the statistical tests to accept or reject the hypotheses are also described in this chapter.

This chapter concludes with a summary and states that the limitations of this study (quality and integrity of the data available for training and testing of the network).

Chapter 5

Research results

5.1. Introduction

The main aim of this chapter is to outline the analysis of the research findings. The analysis approach outlined in this chapter is based on statistical methods. In general, research findings can be split into two main categories, qualitative data and quantitative data. The research data obtained was purely quantitative, therefore the analysis in this chapter uses only quantitative methods.

Descriptive statistics are used to explore the samples used for training, validating and testing of the ANN. Inferential statistics are used to investigate relationships and associations between the sample variables and make inferences regarding the population (the total customer credit base). Data preparation and coding are discussed in this chapter.

The chapter concludes with the presentation of the results from the LVQ ANN training, validation and test modes of operation. The results are also analysed and interpreted.

5.2. Descriptive statistics

The demographic, biographic and behavioural variables (Gender, Race Group, Language and Age Group, etc) frequency representation is important as it can identify areas of interest. Descriptive statistics present an initial “feel” for the data.

The total sample of 2000 customer records is divided into three sub-samples – the training dataset, the validation dataset and the testing dataset. The training dataset will be used to train the ANN such that it may generalize when presented with data that it has not previously encountered. The validation set of 200 records is used to choose the optimal ANN structure for generalization (e.g. number of hidden layer neurons and optimal number of training iterations). The test dataset consists of 400 customer records.

Table 6, below, shows the absolute frequency distribution of key variables of the total sample of 2000 customer records. Table 7 shows the total sample percentage distribution and Table 8 shows the summary statistics of the total customer record sample. The data code book used for the variables in these tables was shown previously in Table 5.

Table 6 - Sample frequency distribution

FREQUENCY DISTRIBUTION													
Variable Type	**Nominal**	**Ordinal**	**Nominal**	**Nominal**	**Nominal**	**Nominal**	**Ordinal**	**Ordinal**	**Ordinal**	**Ordinal**	**Ordinal**	**Ordinal**	**Ordinal**
Variable Name	**Gender**	**Age**	**Race**	**Marital Status**	**Employ-ment Type**	**Residential Status**	**Annual Income**	**Outstanding Balance**	**Ceil On Short Term Loan**	**Time with bank**	**National ity**	**Pref. Lang of Comm**	**Risk Grade**
1	1053	20	447	606	890	574	1500	1171	1601	144	1801	1707	400
2	947	361	455	203	61	1426	60	290	62	236	19	293	400
3		683	85	914	44		43	133	41	160	157		400
4		499	96	99	927		39	157	59	128	23		400
5		250	917	59	20		37	72	49	105			400
6		109		7	58		61	53	34	93			
7		60		112			53	32	36	69			
8		18					36	22	24	36			
9							94	17	20	1029			
10							46	20	19				
11							8	5	55				
12							4	14					
13							11	14					
14							8						
TOTAL	**2000**	**2000**	**2000**	**2000**	**2000**	**2000**	**2000**	**2000**	**2000**	**2000**	**2000**	**2000**	**2000**

Table 7 - Sample percentage distribution

PERCENTAGE DISTRIBUTION													
Variable Name	Gender	Age	Race	Marital Status	Employment Type	Residential Status	Annual Income	Outstanding Balance	Ceil On Short Term Loan	Time with bank	Nationality	Pref Lang of Comm	Risk Grade
1	52.65	1.00	22.35	30.30	44.50	28.70	75.00	58.55	80.05	7.20	90.05	85.35	20.00
2	47.35	18.05	22.75	10.15	3.05	71.30	3.00	14.50	3.10	11.80	0.95	14.65	20.00
3		34.15	4.25	45.70	2.20		2.15	6.65	2.05	8.00	7.85		20.00
4		24.95	4.80	4.95	46.35		1.95	7.85	2.95	6.40	1.15		20.00
5		12.50	45.85	2.95	1.00		1.85	3.60	2.45	5.25			20.00
6		5.45		0.35	2.90		3.05	2.65	1.70	4.65			
7		3.00		5.60			2.65	1.60	1.80	3.45			
8		0.90					1.80	1.10	1.20	1.80			
9							4.70	0.85	1.00	51.45			
10							2.30	1.00	0.95				
11							0.40	0.25	2.75				
12							0.20	0.70					
13							0.55	0.70					
14							0.40						
15													
TOTAL	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Table 8 - Sample summary statistics

TYPE N = Nominal O = Ordinal REVERSE SCORED	N	O	N	N No	N No	N No	O No	O No	O No	O No	O No	O No	O No
VARIABLE	**Gender**	**Age**	**Race**	**Marital Status**	**Employ-ment Type**	**Resident-al Status**	**Annual Income**	**Outstand-ing Balance**	**Ceiling On Short Term Loan**	**Time with bank**	**Nationality**	**Pref. Lang of Comm**	**Risk Grade**
n	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000
Mean		3.63					2.41	2.31	1.98	6.36	1.20	1.15	3.00
Standard Deviation		1.33					2.86	2.29	2.38	3.05	0.62	0.35	1.41
Minimum		1.00					1.00	1.00	1.00	1.00	1.00	1.00	1.00
Q1		3.00					1.00	1.00	1.00	3.00	1.00	1.00	2.00
Median		3.00					1.00	1.00	1.00	9.00	1.00	1.00	3.00
Q3		4.00					1.25	3.00	1.00	9.00	1.00	1.00	4.00
Maximum	2.00	8.00	5.00	7.00	6.00	2.00	14.00	13.00	11.00	9.00	4.00	2.00	5.00
Skewness		0.82					1.97	2.39	2.60	-0.54	2.94	2.00	0.00
Kurtosis		0.55					2.74	6.12	5.86	-1.39	7.28	2.01	-1.30
Mode	1.00	3.00	5.00	3.00	4.00	2.00	1.00	1.00	1.00	9.00	1.00	1.00	1.00

The data in the tables above reflects the following information about the demographic and biographic variables:

Gender: There are 1053 males and 947 females in the sample resulting in a percentage distribution of approximately 53% males and 47% females and a mode (most frequently occurring value) of 1. Males and females are approximately equally represented by the sample.

Age category: The most frequently occurring age group in the sample is 3 (31-40 year old), followed by 4 (41-50 year old) and 2 (21-30 year old). The summary statistics show the first quartile (Q1) to be 3 and the top quartile (Q3) to be 4. This indicates that 50% of the sample lies between age groups 3 and 4 (31-50 years old). These are indications of the variability of the age group (inter-quartile range removes extreme values – doesn't use all the data). Skewness, which is a measure of symmetry of the distribution, is 0.82. Kurtosis, the "flatness" or "peakedness" is 0.55, therefore approximately mesokurtic (average peak). Figure 13 shows the distribution of the age categories with a slight positive skew.

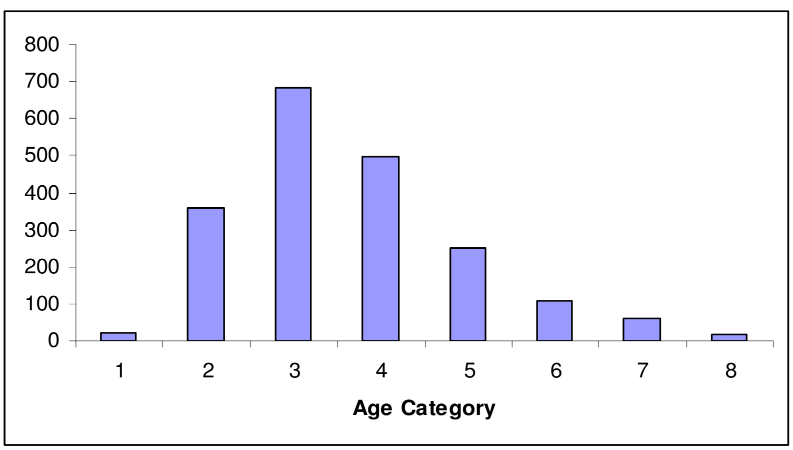


Figure 13 - Age group distribution

Race group: The distribution indicates approximately 22% of the sample is category 1 (White), 23% of the sample is category 2 (Black), 4% is category 3 (Coloured), 5% are category 4 (Asian) and a significant 46% has missing data (shown in Figure 14).

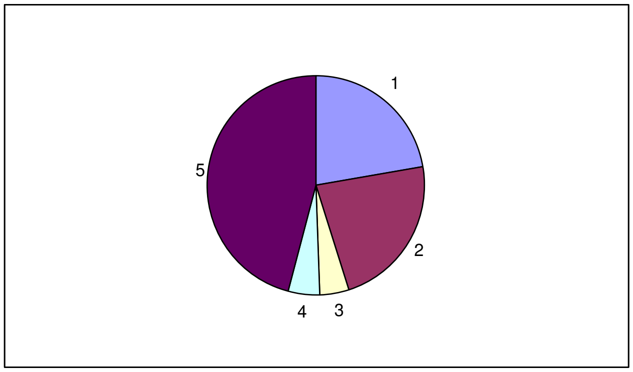


Figure 14 - Race group distribution

Marital Status: The distribution indicates that the majority of customers are single (46%), followed by those married in community of property (30%), those married out of community of property (10%), divorced (5%), widowed (3%), tribal marriages (0.4%) and 6% with missing data or "unknown" status.

Employment Type: The majority of the data (46%) for employment category of customers is classified as "Employed (No type/ unknown)". Customers with professional careers comprise forty five percent of the total sample. Three percent of the sample is self-employed, a further three percent are unemployed, two percent have retired from their jobs and one percent of the sample are housewives.

Residential status: Twenty nine percent of the sample are home owners, while there is unknown data the remaining seventy one percent of the sample.

5.3. Comparing sample demographic groupings and credit risk categories

One of the objectives of this research is to investigate if there is an association between demographic, biographic and behavioural characteristics of the customer and their level of credit risk. This section investigates relationships between the demographic characteristics and the CRG of the customers in the total sample data.

5.3.1. Relationship between gender and credit risk ratings

The Mann-Whitney *U* test is used to test for a relationship between gender and the credit risk grade ratings assigned by the bank. This test is appropriate since two groups (male and female) are being compared. The results of the Mann-Whitney *U* test are shown in Table 9 below.

Table 9 - Relationship between gender and credit risk

Test	**Mann-Whitney test**			
	Credit Risk Analysis			
Alternative hypothesis	Risk Grade by Gender: F ≠ M			
Performed by	Viresh Moonasar			
n	2000			
Risk Grade by Gender	**n**	**Rank sum**	**Mean rank**	**U**
F	947	923873.5	975.58	522195.5
M	1053	1077126.5	1022.91	474995.5
Difference between medians	-			
95% CI	-	to -		
Mann-Whitney U statistic	522195.5			
2-tailed p	0.0618	(normal approximation, corrected for ties)		

The mean rank, above, is the sum of the ranks for each group divided by the number of cases. The closeness of the ranks between the 2 groups is an indication that there is little difference between the means of the 2 groups.

The p-value is not significant and the null hypothesis that the CRG ratings are equal between males and females cannot be rejected. The CRG scores are therefore not significantly different for males and females. It seems unlikely that gender will be a differentiating factor that can assist the ANN to categorize credit risk. Inter-dependencies with other variables may be useful for categorising.

5.3.2. Relationship between age group categories and credit risk ratings

The Kruskal-Wallis one-way analysis of variance test is used to compare the relationship between credit risk ratings and the age groups (both have an ordinal level of measurement). This test is used instead of the Mann-Whitney above since there are more than two (*k*) independent age group categories.

Table 10, below, shows a significant *p*-value (< 0.0001). The null hypothesis that there is no relationship between the age group categories and credit risk ratings is therefore rejected. There is a relationship between age and credit risk categories of customers. Age is therefore used as an element of the input vector that will be fed to the ANN credit risk model.

Table 10 – Relationship between age categories and credit risk

Test	**Kruskal-Wallis ANOVA**		
	Credit Risk Analysis		
Comparison	Risk Grade by Age: 1, 2, 3, 4, 5, 6, 7, 8		
Performed by	Viresh Moonasar		
n	2000		
Risk Grade by Age	**n**	**Rank sum**	**Mean rank**
1	20	8410.0	420.50
2	361	449580.5	1245.38
3	683	743741.5	1088.93
4	499	525249.5	1052.60
5	250	206525.0	826.10
6	109	40654.5	372.98
7	60	22430.0	373.83
8	18	4409.0	244.94
Kruskal-Wallis statistic	373.05		
p	<0.0001	(chisqr approximation, corrected for ties)	

5.3.3. Relationship between race group and credit risk ratings

The results of the Kruskal-Wallis one-way analysis of variance test that tests for a relationship between race groups and their credit risk ratings are shown in Table 11, below. The significant *p*-value (< 0.0001) forces the rejection of the null hypothesis that there is no difference between the credit risk ratings amongst the different race groups.

Table 11 - Relationship between race group and credit risk

Test	**Kruskal-Wallis ANOVA**		
Comparison	Credit Risk Analysis Risk Grade by Race: 1, 2, 3, 4, 5		
Performed by	Viresh Moonasar		
n	2000		
Risk Grade by Race	n	Rank sum	Mean rank
1	447	380023.5	850.16
2	455	564827.5	1241.38
3	85	83442.5	981.68
4	96	109648.0	1142.17
5	917	863058.5	941.18
Kruskal-Wallis statistic	130.21		
p	<0.0001	(chisqr approximation, corrected for ties)	

5.3.4. Relationship between marital status and credit risk ratings

The significant *p*-value (< 0.0001), shown below in Table 12, forces the rejection of the null hypothesis that there is no difference between the credit risk ratings amongst the customers with different marital status.

Table 12 - Relationship between marital status and credit risk

Test	**Kruskal-Wallis ANOVA**		
Comparison	Credit Risk Analysis Risk Grade by Marital Status: 1, 2, 3, 4, 5, 6, 7		
Performed by	Viresh Moonasar		
n	2000		
Risk Grade by Marital Status	**n**	**Rank sum**	**Mean rank**
1	606	600703.0	991.26
2	203	143101.5	704.93
3	914	1019657.0	1115.60
4	99	86649.5	875.25
5	59	42629.5	722.53
6	7	7803.5	1114.79
7	112	100456.0	896.93
Kruskal-Wallis statistic	116.50		
p	<0.0001	(chisqr approximation, corrected for ties)	

5.3.5. Relationship between employment type and credit risk ratings

The significant *p*-value (< 0.0001), shown below in Table 13, forces the rejection of the null hypothesis that there is no difference between the credit risk ratings amongst the customers with different types of employment.

Table 13 - Relationship between employment type and credit risk

Test	**Kruskal-Wallis ANOVA**		
Comparison	Credit Risk Analysis Risk Grade by Employment Type: 1, 2, 3, 4, 5, 6, 7		
Performed by	Viresh Moonasar		
n	2000		
Risk Grade by Employment Type	**n**	**Rank sum**	**Mean rank**
1	890	896845.0	1007.69
2	61	55830.5	915.25
3	44	13622.0	309.59
4	231	278715.5	1206.56
5	20	10410.0	520.50
6	58	76429.0	1317.74
7	696	669148.0	961.42
Kruskal-Wallis statistic	133.71		
p	<0.0001	(chisqr approximation, corrected for ties)	

5.3.6. Relationship between residential status and credit risk ratings

No statistical analysis was conducted as majority of the data is missing or unknown. This could result in results that are not meaningful and was therefore omitted.

5.4. Assessing the strengths of association

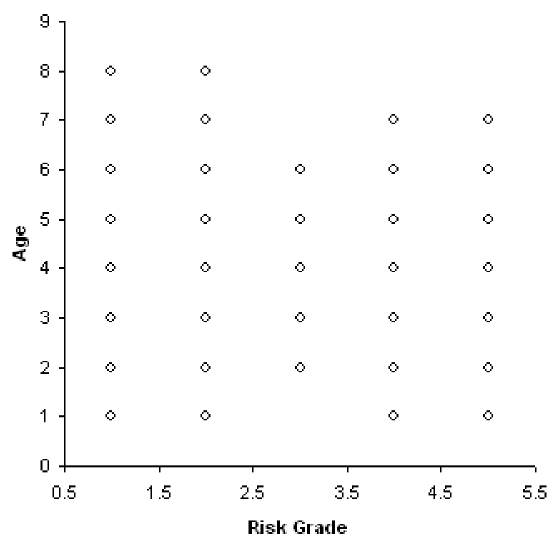
The previous section tested if relationships existed between the various demographic characteristics and credit risk ratings. This section evaluates the strengths and direction (if applicable) of association between the characteristic and the credit risk rating.

5.4.1. Strength of association between age group and credit risk ratings

Spearman rank correlation test is used to measure the strength and direction of association between the age group of the customer and the CRG (both variables concerned are ordinal). The Spearman's rank-order correlation coefficient (rs) statistic in Table 14 shows that there is a moderate, negative association between the age group and the CRG i.e. younger customers are associated with greater risk. The rs coefficient has a range from -1 to +1 and indicates the strength of association.

Table 14 - Strength of association: Age and credit risk grade

Test	**Spearman rank correlation**	
	Credit Risk Analysis	
Alternative hypothesis	Risk Grade ≠ Age	
Performed by	Viresh Moonasar	
n	2000	
rs statistic	-0.33	
95% CI	-0.37	to -0.29
2-tailed p	<0.0001	(t approximation, corrected for ties)



This input characteristic therefore has the potential of assisting the ANN model to determine the credit risk level of a customer.

5.4.2. Strength of association between marital status and credit risk ratings

Spearman's rank-order correlation cannot be used to test the strength of association between the marital status and the credit risk ratings since the level of measurement of marital status is nominal. Instead, Cramer's *V* test is used to measure the strength of association between these two variables. Cramer's *V* has values that range from 0 to 1 and cannot determine the direction of association because of the nature of the nominal level of measurement. Table 15, below, shows the Chi-square statistic that is used to derive the value of Cramer's *V*.

Table 15 - Strength of association: marital status and credit rating

Test	**Chi-square test**							
	Credit Risk Analysis							
Comparison	Risk Grade by Marital Status							
Performed by	Viresh Moonasar							
n	2000							
	Marital Status							
Risk Grade	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Total**
1	138 (121.2)	93 (40.6)	89 (182.8)	30 (19.8)	27 (11.8)	0 (1.4)	23 (22.4)	400
2	92 (121.2)	36 (40.6)	203 (182.8)	20 (19.8)	9 (11.8)	2 (1.4)	38 (22.4)	400
3	139 (121.2)	19 (40.6)	196 (182.8)	16 (19.8)	9 (11.8)	2 (1.4)	19 (22.4)	400
4	120 (121.2)	38 (40.6)	208 (182.8)	17 (19.8)	6 (11.8)	2 (1.4)	9 (22.4)	400
5	117 (121.2)	17 (40.6)	218 (182.8)	16 (19.8)	8 (11.8)	1 (1.4)	23 (22.4)	400
Total	606	203	914	99	59	7	112	2000
X^2 statistic	221.04							
p	<0.0001							

$$\text{Cramers } V = \sqrt{\frac{\chi 2}{n \quad (k-1)}}$$

n is the sample size, k is the smaller of the number of rows or columns in the contingency table.

$$\text{Cramers } V = \sqrt{\frac{221.04}{2000 \quad (5-1)}} \;=\; 0.17$$

There is a low level of association between marital status and level of credit risk.

5.4.3. Strength of association between race group and credit risk ratings

Table 16, below, shows the Chi-square statistic that is used to derive the value of Cramer's *V* that indicates the strength of association between race groups and their credit ratings.

Table 16 - Strength of association: race group and credit rating

Test	**Chi-square test**
	Credit Risk Analysis
Comparison	Risk Grade by Race
	Viresh Moonasar
Performed by	
n	2000

	Race					
Risk Grade	**1**	**2**	**3**	**4**	**5**	**Total**
1	143 (89.4)	14 (91.0)	11 (17.0)	6 (19.2)	226 (183.4)	400
2	105 (89.4)	70 (91.0)	30 (17.0)	16 (19.2)	179 (183.4)	400
3	61 (89.4)	155 (91.0)	12 (17.0)	36 (19.2)	136 (183.4)	400
4	53 (89.4)	60 (91.0)	16 (17.0)	14 (19.2)	257 (183.4)	400
5	85 (89.4)	156 (91.0)	16 (17.0)	24 (19.2)	119 (183.4)	400
Total	447	455	85	96	917	2000

X^2 statistic	345.88
p	<0.0001

$$\text{Cramers } V = \sqrt{\frac{345.88}{2000 \quad (5-1)}} \;=\; 0.21$$

There is a low level of association between race group and level of credit risk.

The results of comparing sample demographic groupings and credit risk categories have shown while there is no difference between gender and credit risk ratings, there were significant differences between age groups, race groups, customer's marital status and employment type with credit risk ratings. These differences could be used to train an ANN to differentiate between customers of different risk categories.

The next section describes the results obtained when using the characteristic data (demographic, biographic and behavioural) to train and test the ANN LVQ credit model.

5.5. LVQ training mode results

All training, validation and testing of the LVQ neural network is conducted in the MATLAB software environment and implemented on a standard personal computer. The specification of the system that is used is shown in Table 17 below.

Table 17 - PC specifications

CPU Type	Intel Centrino Duo T2400
CPU Speed	1.83 GHz
Random Access Memory (RAM)	1 Gigabyte
Operating System	Microsoft XP (with Service Pack 2)
Simulation Environment	MATLAB by MATHWORKS

The input feature space was constructed by importing the input variables from a comma separated variable (CSV) formatted file into the MATLAB environment. The corresponding LVQ output target vector was also similarly imported into MATLAB. The LVQ was initialized and then trained with the imported datasets. The training process continually adjusts the weight matrices with each iteration of input and target vector pairs. The MATLAB program code used is shown in Appendix D.

Figures 15 and 16 show the LVQ training progress of the LVQ network. The network shown typically imports all training and target vector sets into MATLAB and trains a LVQ network (80 hidden layer neurons) with the specified 5000 epochs in approximately 15 seconds (with the plot shown in Figure 16 running in the background). Training time is increased to 127 seconds if the plot is displayed and updated during training.

Figure 15 - LVQ training progress

Figure 16 is a plot of the input vectors labelled according to their target vectors (low risk: o, high risk: X). The first part shows the distance between the input vectors and their targets after just 1 epoch or training cycle and the second part shows the training and target pairs after the maximum specified 5000 epochs.

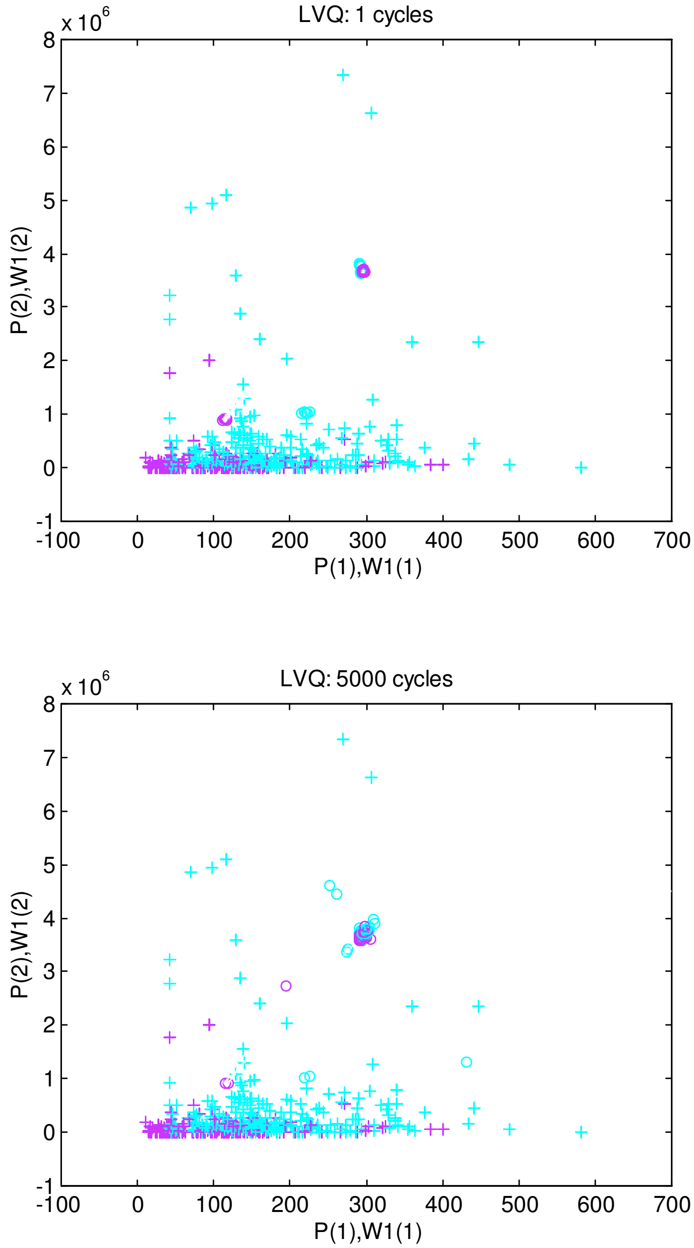


Figure 16 - LVQ training neuron updates

5.6. LVQ validation mode results

The validation data set is used to choose the optimal LVQ network parameters to avoid problems that are commonly associated with neural networks: under-fitting, over-fitting, incorrect learning rate and excessive training. Learning rate (lr) was held constant at 0.01 and the maximum number of epochs (me) was 5000 for all networks. Table 18 shows the different architectures used for validation.

Table 18 - Network architecture and specifications

	Number of hidden layer neurons
Network 1	20
Network 2	40
Network 3	60
Network 4	80
Network 5	100
Network 6	120

Table 19 shows the results obtained by each LVQ network. Network 5 has the greatest Type II accuracy at the expense of an inferior Type I accuracy.

Table 19 - Validation mode test results

	Type I accuracy (% bad)	**Type II accuracy (% good)**	**Total accuracy (%)**
Network 1	72	72	72
Network 2	82	58	70
Network 3	82	48	65
Network 4	72	72	72
Network 5	**52**	**88**	**70**
Network 6	74	72	73

5.6.1. Analysis of the validation mode results

The Type II accuracy indicates the ability of the ANN to correctly classify good customer credit records. It is important that the Type II accuracy is maintained at a high level as the financial institute would prefer to rather classify a bad customer incorrectly rather than misclassify a good customer and lose revenue generating income. Since all current customers have been previously classified as good credit risk cases by the institutes current risk assessment system (normally manually analysed by financial experts), any improvement in the classification of bad customer records can result in possible prevention of revenue loss through early automatic detection of bad customer records. Network 5 is therefore selected for the test mode analysis of the LVQ ANN. Networks 1, 2, 3 and 4 exhibit symptoms of under-fitting whilst network 6 may suffer from over-fitting and poor generalizing capability.

5.7. LVQ test mode results

The final values of the ANN weight matrices are determined upon completion of the training mode. The network is then presented with the test dataset to evaluate its performance in classifying the credit risk class of each customer. The ability of the network to generalize is being evaluated in this phase as the ANN has not being previously exposed to the test dataset during training and validation. The test mode dataset consists of 400 customer records (200 “good” and 200 “bad” customer records).

The results from the test mode are presented in Table 20 and Table 21 below. Test results for Network 5 are optimal as predicted previously by the validation results. The results obtained with the other network architectures are included as a point of interest.

Table 20 - Test mode data results

	Total “bad” customer records	Number of “bad” records incorrectly classified as “good”	Type I accuracy (classified correctly “bad” by LVQ ANN)	Total “good” customer records	Number of “good” records incorrectly classified as “bad”	Type II accuracy (classified correctly “good” by LVQ ANN)	Total number classified accurately by LVQ ANN
Network 1	200	34	166	200	88	112	278
Network 2	200	64	136	200	56	144	280
Network 3	200	34	166	200	88	112	278
Network 4	200	71	129	200	43	157	276
Network 5	**200**	**92**	**108**	**200**	**32**	**168**	**276**
Network 6	200	21	179	200	112	88	267

Table 21 - Test mode results (%)

	Type I accuracy (% bad)	Type II accuracy (% good)	Total accuracy (%)
Network 1	83	56	70.5
Network 2	78	72	70.0
Network 3	83	56	70.5
Network 4	64.5	78.5	69.0
Network 5	**54**	**84**	**69.0**
Network 6	89.5	44	66.7

5.7.1. Analysis of LVQ test mode results

Network 5, with 100 hidden layer neurons, has a Type I accuracy of 54%, a Type II accuracy of 84% and a total credit risk classification accuracy of 69%. This network correctly classifies 54% of "bad" customer credit risks and 84% of current "good" customer credit risks. There seems to be potential savings by early, automatic detection of the "bad" credit risks which can be prioritized and analyzed further by financial experts.

Should the output of the ANN not be able to significantly correlate with the actual credit performance of the consumers or produce a *rho* correlation value less than that obtained with conventional methods then the contribution made to the body of knowledge could be that further data pre-processing and/or additional data on consumer attributes is required in order to find relationships between the consumer characteristic data and their credit behaviour.

5.7.2. Comparison of results with previous research

The results obtained in this research are comparable to that achieved by Komo'rad (2002) and Jagielska and Jaworski (1996). In the latter work, the following results were achieved by an ANN BNN network with two hidden layers and ten neurons:
Type II accuracy of 83.7%, Type I accuracy of 15.0 and a total accuracy of 70%. The LVQ network used in this research performs with greater Type I and Type II accuracy. Type II accuracy is increased to 54%.

5.8. Summary

This chapter presented the results of the various statistical tests on the sample data and the ANN LVQ credit risk model. Descriptive and summary statistics were used to explore and present an initial “feel” for the data.

The Mann-Whitney U and Kruskal-Wallis one way ANOVA tests were used to test for relationships between characteristic input variables and the credit risk rating of customers. While the null hypothesis that there is no relationship between gender and credit risk grades could not be rejected, significant relationships were found between age groups, race groups, marital status and employment types of customers and their credit risk grades.

The performance of the LVQ ANN was tested on various network architectures (by varying the number of hidden layer neurons) on the validation dataset. Network 5, with 100 hidden layer neurons, performed with the highest Type II accuracy and was chosen as the optimal architecture for evaluation on the test dataset. Network 5 obtained a Type II accuracy of 84%, Type I accuracy of 54% and an overall accuracy of 69% when tested on the test dataset.

Chapter 6

Discussions, conclusion and recommendations

6.1. Introduction

The goal of this research report was to present the findings of the investigation, communicate the recommendations and conclusion and suggest areas of further research. Therefore the first section of this chapter provides a summary of the research findings. This includes all the achievements accomplished by conducting this research. The second section of this chapter outlines the recommendations and conclusion. The aim in this section is to prove that the suggested recommendations and conclusion are logically derived from the analysis of the findings. The last section of this chapter is a list of suggestions for further research.

6.2. Summary of the research findings

It is important to note that the research objectives have been achieved. The main research problem was to test whether LVQ artificial neural networks can be applied accurately to South African consumer credit risk analysis. Sub-problems included investigating if relationships between biographic and demographic characteristics of consumers and their credit risk scores existed, and investigating the effect of varying the ANN network architecture on it ability to detect credit risk at the customer level.

The results of the research have shown that LVQ ANN may be used to detect and predict customer risk grade scores. Hypotheses testing have been utilized to show that relationships exist between age groups, marital status, race groups, employment type and the credit risk grades of customers. LVQ ANN architecture, learning rates and number of training epochs have also been found to affect the performance of the recognition system. The Type II accuracy of the system indicates the potential to identify “bad” credit risks early thereby reducing loss of revenue.

6.3. Recommendations

After close examination and analysis of the research findings, the following recommendations are suggested. The recommendations are suggested based on the research hypotheses.

Plan for credit risk analysis:
The data utilized has not been collected, by the financial institution, specifically for the purpose of credit risk analysis. The data has been extracted from several individual databases and does contain much missing information. Accuracy and integrity of data is very important for future studies of credit risk analysis. It is therefore recommended that data is collected from each customer that contains all the fields necessary for credit risk analysis. However, this may result in increased cost (both time and money). It is a trade-off between expensive data and between low accuracy due to not enough information.

Structure data and information systems appropriately:
Unambiguous data can be minimised by designing databases and information systems (IS) that do have the capacity of allowing for missing or unavailable data e.g. if there is no record of a customers annual salary then there should be a value that can uniquely represent that instead representing it with the value “0” which confuses missing data with those customers who currently have no income. It is essential that an Information strategy is in place that can be supported by appropriate IS.

Enable credit data validation:
Inaccuracies in credit reports can result in serious disadvantages for consumers due to the importance of credit reports. Customers should be allowed to verify their records and financial institutions to amend inconsistencies after verifying claims about incorrect data. The Financial Information and Intelligence Credit Act (FICA) is a step in the right direction.

6.4. Suggestions for further research

The dynamic aspects and structural changes over time are not considered in this study as the data to be obtained is not conducive to a longitudinal study. It is important to investigate the progress of those “good” accounts that were incorrectly classified (Type II error). The LVQ could have predictive capability that may result in early identification of problematic accounts.

Further research is needed particularly to incorporate the relative costs of rejected good applicants and accepted slow payers, using larger datasets which preserve the proportions of good and slow payers in the population.

Another key aspect of research that may investigate the genuine predictive nature of credit risk analysis models is using only data at time of credit application and tracking progress of applicant against classification of credit risk model for a specified duration of time.

Reliability and validity of the current work can be researched further. The accuracy, meaningfulness, and credibility of the research project could be enhanced through the use of triangulation. Data could be separately analyzed from multiple sources.

6.5. Conclusion

In conclusion it was observed that credit scoring represents a set of common techniques to decide whether a bank or financial institution should grant a loan (issue a credit card) to an applicant or not. Relationships do exist between a customer’s characteristic demographic, biographic and behavioural variables and his/her credit risk grading. Artificial neural networks are very flexible models that may be used as a decision support system to automatically identify and prioritize “bad” customer accounts. The consistency of unbiased evaluation (free of human error and emotion) can be ensured.

Although ANN models have been found to successfully predict credit risk grades, it in no way suggests substitution of experts by the software model. Rather the ANN model rating can be used as a bench mark indication to complement other methods e.g. financial expert/ consultant analysis. Therefore, an ANN model can successfully be used as a Decision Support System.

List of References

Atiya, F. A., 2001. ‘Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results’. *IEEE TRANSACTIONS ON NEURAL NETWORKS,* 12(4): 929-935.

Bandyopadhyay, A. 2006. ‘Predicting probability of default of Indian corporate bonds: logistic and Z-score model approaches’. *The Journal of Risk Finance*, 7(3): 255-272

Clayton, C. 2005. *All-encompassing credit bill set to shake up market.*
[online] Personal Finance.
Available from
<<http://www.persfin.co.za/index.php?fSectionId=588&fArticleId=2528370>>
[Accessed 1 May 2007]

Cooper, D. R. and Schindler, P. S. no date. *Business Research.* [online]
Available from <<http://www.c-r-g.com/busresearch/>>
[Accessed 1 May 2007]

Crook, J. N., Thomas, *L. C.* & Hamilton, R. 1994. ‘Credit Cards: Haves, Have-Nots and Cannot-Haves’. *Service Industries Journal*, 14(2): 204-215.

Demuth, H. and Beale, M. 1994. *Matlab Neural Network Toolbox - User’s guide,* Natick: USA.

Diamantopoulos, A. and Schlegelmilch, B.B. 2005. *Taking the Fear Out of Data Analysis*, Thompson Learning, Singapore.

Diana, T. 2005. *Credit Risk Analysis and Credit Scoring – Now And In The Future*
[online] National Association of Credit Management.
Available from <<http://www.ecredit.com/docs/pdf/TOM_MARCH.pdf>>
[Accessed 17 May 2007]

(Electronic Privacy Information Centre, 2003). *Credit Scoring*.
[online] Electronic Privacy Information Centre

Available from <http://www.epic.org/privacy/creditscoring/#Problems>
Accessed on 29 May 2007.

Eurofinas. 2005. 'Consumer Credit, Car Financing, Industrial Credit & Real Estate Mortgage Activity in Europe', *Eurofinas Statistical Bulletin 9: 1-15*

Gunnion, S. 2007. *South Africa: Banks Move to Sort Out Credit 'Madness' [online]*. Johannesburg: Business Day.
Available from:
<<http://allafrica.com/sendpage.html?ref=http://allafrica.com/stories/200703020173.html>>
[Accessed on: 5 March 2007]

Haykin, S. 1994. *Neural Networks: A Comprehensive Foundation,* Macmillan, NY.

Ho, Y. C. and Pepyne, D. L. 2002. 'Simple Explanation of the No-Free-Lunch Theorem and its Implications'. *Journal of Optimization Theory and Applications*, 115(): 549.

Huang, Z., Chen, H., Hsu, C., Chen, W. & Wu, S. 2004. 'Credit rating analysis with support vector machines and neural networks: a market Comparative study'. *Decision Support Systems.* 37:543–558.

Jagielska, I. and Jaworski, J. 1996. 'Neural Network for Predicting the Performance of Credit Card Accounts'. *Computational Economics,* 9(): 77-82.

Kohonen, T. 1997. '*Self-Organizing Maps', 2nd Edition,* Springer-Verlag, *Berlin*.

Komor´ad, K. 2002. On Credit Scoring Estimation. Master of Science thesis. Institute for Statistics and Econometrics, Humboldt University, Berlin.

Kohonen, T. 1997. *Self-Organizing Maps.* 2nd ed, Berlin: Springer-Verlag.

Krauss, T. P. 1995. *Signal Processing Toolbox*, The MathWorks Inc.

Krishnaswamy, C. R., Gilbert, E. W. & Pashley, M. M. 2000. 'Neural Network Applications in Finance: A Practical Introduction'. *Financial Practice & Education*, 10(1): 75-84.

Kumar, K. and Haynes, J. D. 2003. 'Forecasting Credit Ratings Using ANN and Statistical Techniques'. *International Journal of Business Studies,* 11(1): 91-108.

Länsiluoto, A., Eklund, T, Back, B. Vanharanta, H. & Visa, A. 2004. 'Industry-specific cycles and companies' financial performance comparison using self-organizing maps'.
Benchmarking: An International Journal, 11(3): 267-286.

Lai, K. K. Yu, L. Wang, S. & Zhou, L. 2006a. 'Credit Risk Analysis Using a Reliability-Based Neural Network Ensemble Model'**.** In**:** Kollias, S. ed. *ICANN 2006, Part II, LNCS 4132.* Springer-Verlag: Berlin. 682 – 690.

Lai, K. K. Yu, L. Wang, S. & Zhou, L. 2006b. 'Neural Network Metalearning for Credit Scoring'. In: Huang, D.S. Li, K. and Irwin, G. W. eds. *ICIC 2006, LNCS 4113.* Springer-Verlag: Berlin. 403 – 408.

Lanquillon, C. 1999. 'Dynamic Aspects in Neural Classification', *International Journal of Intelligent Systems in Accounting Finance & Management,* 8(4): 281-296.

Leedy, P. D. and Ormrod, J. E. 2005. *Practical Research*. New Jersey: Pearson - Merill Prentice Hall.

Li, A., Shi, Y., Zhu, M. & Dai, J. *A Data Mining Approach to Classify Credit Cardholders: proceedings of IEEE International Conference on Data Mining Workshops (ICDMW'06), December,* 2006. 828-832

Lorenzo, J. M. M., Martinez, M. G. and Garcia, N. 2000. 'Rating Small and Medium-Sized Firms for Granting Credit Using Artificial Neural Networks', *International Advances in Economic Research*, 6 (4): 794.

Mester, L. J. 1997. 'What's the Point of Credit Scoring?', Business Review, Federal Reserve Bank Of Philadelphia : 1-16.

Moonasar, V. and Venayagamoorthy, G.K. 2001. 'Automatic Speaker Recognition (ASR) Systems using Neural Networks: A Committee Approach', *Journal of the School of Postgraduate Studies, ML Sultan Technikon*, 2(1):10-17.

Ripley, B. D. (1996) *Pattern Recognition and Neural Networks,* Cambridge: Cambridge University Press.

Sarle, W. S. 2002. NNET FAQS [online],
Avaliable from <ftp://ftp.sas.com/pub/neural/FAQ.html>
Accessed on 8 May 2007

Scarborough, D. and Somers, M. J. 2006. *Neural Networks in Organizational Research: Applying Pattern Recognition to the Analysis of Organisational Behavior*. Washington, DC: American Psychological Association.

Sherwell, P. 2007. 'America dream sours as housing market collapses', Denver: Sunday Telegraph [online],
Available from
http://www.telegraph.co.uk/news/main.jhtml;jsessionid=VDTO2A0QR3HUHQFIQMGCFFWAVCBQUIV0?xml=/news/2007/05/06/whouse06.xml
Accessed on 28 August 2007

The Credit Research Foundation. 2006. 'Credit Scoring: The Future Of Decisioning In The A/R Process', *Business Credit,* 108 (2): 34.

Thomas, L.C. 2002. 'A Survey of Credit and Behavioral Scoring: Forecasting Financial Risk of Lending to Consumers', *International Journal of Forecasting*, 16(1): 149-172.

Trochim, W. M. K., 2006. 'Probability Sampling', The Research Methods Knowledge Base [online],
Available from http://www.socialresearchmethods.net/kb/sampprob.php

Accessed on 30 August 2007

Yobas, M. B., Crook, J. N. & Ross, P. 2000. 'Credit scoring using neural and evolutionary techniques', *IMA Journal of Mathematics Applied in Business and Industry,* 11(1): 111-125

Yu, L., Lai, K. K. & Wang, S. *Credit Risk Assessment with Least Squares Fuzzy Support Vector Machine: proceedings of IEEE International Conference on Data Mining Workshops (ICDMW'06), December,* 2006. 823-827

Wikipedia. 2007b. *Online algorithm* [online]. US: Wikipedia
Available from:
URL<http://en.wikipedia.org/wiki/Online_algorithm#Online_algorithms>
[17 May 2007]

Wikipedia. 2007b. *Credit score* [online]. US: Wikipedia
Available from: URL<http://en.wikipedia.org/wiki/Credit_score>
[17 May 2007]

Zurada, J.M. 1992. *Introduction To Artificial Neural Systems,* Boston: PWS Publishing Company.

List of Appendices

Appendix A – ANN neuron transfer functions

Hard limit transfer function

The output of the neuron is limited to either 0 or 1 depending on whether or not input, p, exceeds -b/w as illustrated in figure 12.

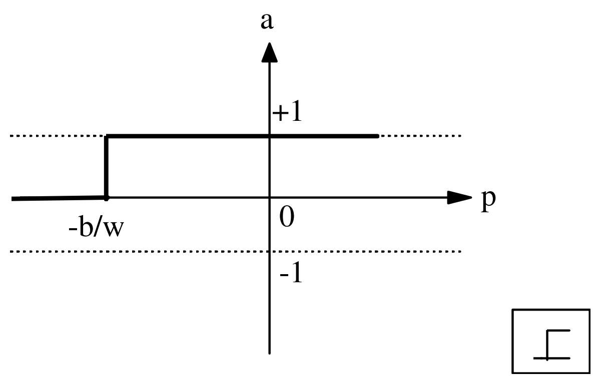


Figure 17 - Hard limit transfer function with bias

The Linear Transfer Function

The input output relationship for a single input neuron with a linear transfer function is shown in figure 13:

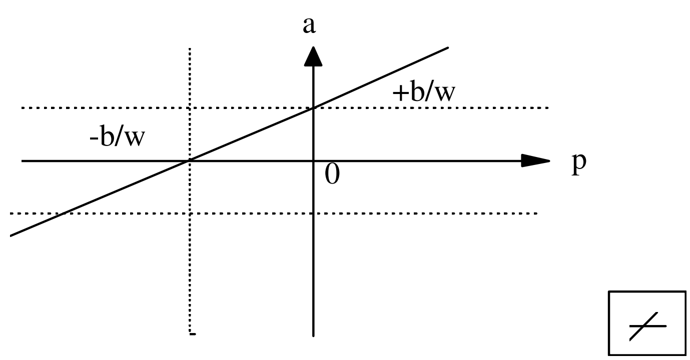


Figure 18 - Linear transfer function

The Sigmoid Transfer Function

The sigmoid transfer function, shown in figure , takes the input (which may have any value between plus and minus infinity) and squashes the output into the range 0 and 1. This transfer function is commonly used in backpropagation networks and is differentiable.

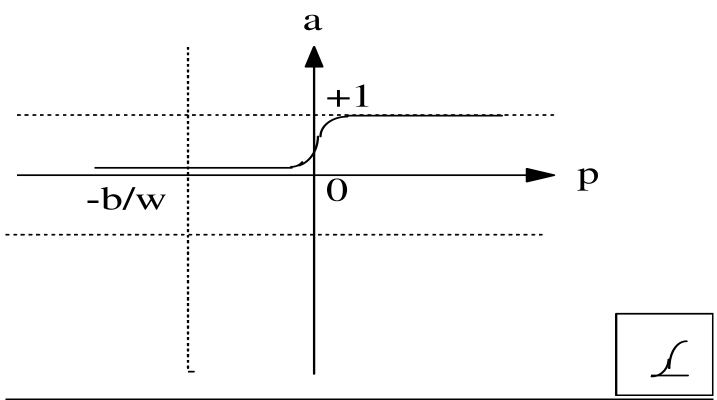


Figure 19 - Sigmoidal transfer function

- The step function

$$f(z) = \begin{cases} 1 : & z \geq 0 \\ 0 : & z < 0 , \end{cases}$$

- The Piecewise-linear function

$$f(z) = \begin{cases} 1 : & z > a \\ z : & -a \leq z \leq a \\ -1 : & z < -a , \end{cases}$$

- The sigmoid function

$$f(z) = \frac{1}{1 + \mathrm{e}^{-z}} ,$$

Appendix B – Back-propagation (BNN)

BNN was created by generalizing the LMS rule multiple layer networks and non-linear differentiable transfer functions. Input vectors and their corresponding output vectors are used to train a network until it can approximate a function, associate input vectors with a specific output vector, or classify input vectors in a way defined by the user. Networks with biases, a sigmoid layer and a linear output layer are capable of approximating any function with a finite number of discontinuities. The generalisation property of BNN makes it possible to train a network on a representative set of input/target pairs and get good results without training the network on all possible input/output pairs.

BNN often use the log-sigmoid transfer function. The function generates outputs between 0 and 1 as the neurons net input goes from negative to positive infinity. Tan-sigmoid and occasionally linear transfer functions may also be used. If the last layer of the network has sigmoid neurons then the outputs of the network are limited to a small range. If linear output neurons are used then the outputs can take on any value. Therefore, BNN networks often have one or more layers of sigmoid neurons followed by a layer of linear neurons. The non-linear transfer function allows the network to learn both linear and non-linear relationships between input and output vectors. The linear output layer lets the network produce values outside the range -1 to +1.

Appendix C - LVQ rules and network architecture

LVQ training requires a training set of examples of the proper network behaviour. If the input pattern is classified correctly, then move the winning weight toward the input vector according to the **Kohonen** rule.

$$_{i^*}W^1(q) = {}_{i^*}W^1(q-1) + \alpha(p(q) - {}_{i^*}W^1(q-1))$$

If the input pattern is classified incorrectly, then move the winning weight away from the input vector according to the rule:

$$_{i^*}W^1(q) = {}_{i^*}W^1(q-1) - \alpha(p(q) - {}_{i^*}W^1(q-1))$$

This is known as LVQ1 training. Other variations exist where the neighbours of the winning weight are adjusted as well as the winning weight (LVQ2 and LVQ3).
The weight vectors of the competitive layer are calculated using the midpoint theorem. The training input vectors must contain expected minimum and maximum values in their range. The net input to the competitive layer is the negative distance between the prototype vectors, W^1 and the input:

$$n_i^1 = -\| {}_iW^1 - p \|$$

Figure 20 below shows an LVQ network with a competitive layer being followed by a linear layer. The classes learned by the competitive layer are termed subclasses and the classes of the linear layer are termed target classes.

The output of the linear layer is given by the relationship:

$$a^2 = W^2 a^1$$

S^1 is the number of neurons in the competitive layer. This parameter is user defined. The magnitude of W^1 is dependent on S^1 according to the relationship:

$$W^1 = S^1 \times R$$

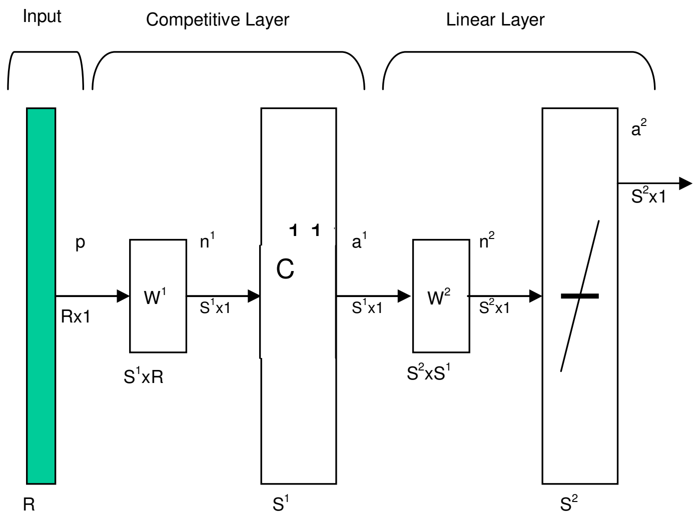


Figure 20 - LVQ network architecture

For the LVQ network, the winning neuron in the first layer indicates the **subclass** which the input vector belongs to. There may be several different neurons (subclasses) which make up each class.

The second layer of the LVQ network combines subclasses into a single class. The columns of $\mathbf{W}^2$ represent subclasses, and the rows represent classes. $\mathbf{W}^2$ has a single 1 in each column, with the other elements set to zero. The row in which the 1 occurs indicates which class the appropriate subclass belongs to.
e.g.

$$\mathbf{W}^2 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Subclasses 1, 3 and 4 belong to class 1.
Subclass 2 belongs to class 2.
Subclasses 5 and 6 belong to class 3.

LVQ2

If the winning neuron in the hidden layer incorrectly classifies the current input, we move its weight vector away from the input vector, as before. However, we also adjust the weights of the closest neuron to the input vector that does classify it properly. The weights for this second neuron should be moved toward the input vector.

When the network correctly classifies an input vector, the weights of only one neuron are moved toward the input vector. However, if the input vector is incorrectly classified, the weights of two neurons are updated, one weight vector is moved away from the input vector, and the other one is moved toward the input vector. The resulting algorithm is called **LVQ2**.

Appendix D – MATLAB sample implementation code of LVQ

```
P=[v1 v2 v3 …vn];                      % P is the input training matrix. Each sample is a column vector consisting of the consumer characteristics.
C=[1 2 3 4 5];                         % output vector with corresponding categories as target output for each input of n column vectors
T=ind2vec(C);                          % plot training progress
colormap(hsv)
plotvec(P,C)
alabel('p(1)','p(2)','input vectors')
S=30;                                  % number of neurons in hidden layer
[W1,W2]=initlvq(P,S,T);                % initialize weights
hold on
plot(W1(1,1),W1(1,2),'ow')             % plot training progress
alabel('p(1),w(1)','p(2),w(3)','input/weight vectors')
df=20;                                 % display frequency
me=10000;                              % maximum epochs
lr=0.001;%load r;                      % learning rate
tp=[df me lr];                         % training parameters
[W1,W2] = TRAINLVQ(W1,W2,P,T,tp)       % train network
%save r
P=v8;
a=simulvq(P,W1,W2)                     % test network
```

Actual code used:

```
P = csvread('credtr16.csv');
P=P'
C=csvread('credta16.csv');
C=C'
T=ind2vec(C);
colormap(hsv)
```

```
plotvec(P,C)
alabel('p(1)','p(2)','input vectors')
S=75;
[W1,W2]=initlvq(P,S,T);
hold on
plot(W1(1,1),W1(1,2),'ow')
alabel('p(1),w(1)','p(2),w(3)','input/weight vectors')
df=20;
me=5000;
lr=0.001;
%load weight.mat
tp=[df me lr];
[W1,W2] = TRAINLVQ(W1,W2,P,T,tp)
Q=csvread('credva13.csv')
Q=Q'
a=simulvq(Q,W1,W2)
```

Appendix E – CRISIL'S Company Rating Methodology

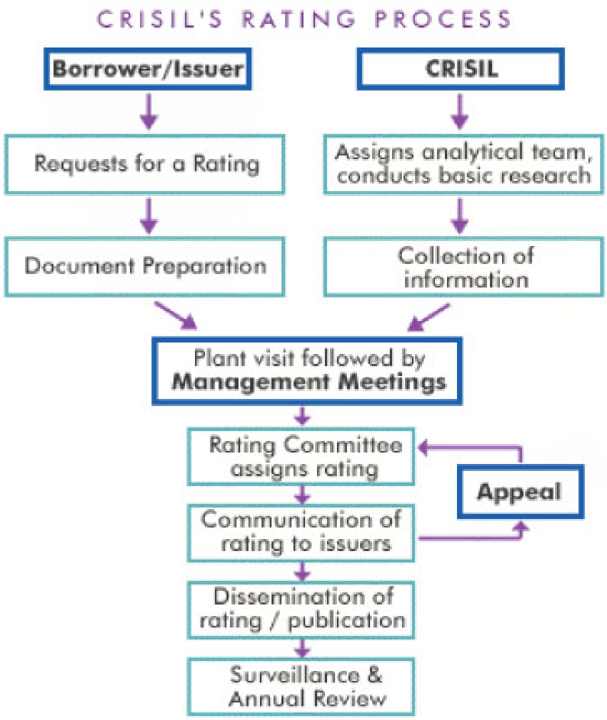


Figure 21 - CRISIL's company credit rating methodology

Figure 16 shows the methodology used to obtain rating reports published in CRISIL rating scan issues of May 00, June 00, July 00 and Sept 00 used in the study by Kumar and Haynes (2003). Either a debenture issue (both convertible and Non-convertible), Short Term instrument or fixed deposit rating is used for rating various classes. Fourteen financial statistics as given by the experts of CRISIL, extracted along with the concomitant rating for each debenture issue/fixed deposit/short term have been taken for 76 companies. These ratios are:

1. Net Sales (NETSALES)
2. Operating Income (OPERINCO)
3. Operating Profit before Depreciation, Interest and Tax (OPBDIT)
4. Profit after Tax (PAT)
5. Equity Share Capital (ECAPITAL)
6. Tangible Networth (NETWORTH)
7. OPBDIT/Operating Income (OPBOPI)
8. PAT/Operating Income (PATOI)
9. PBIT/ (Total Debt + Tangible Networth) - (PBITTDTN)
10. OPBDIT/Interest and Finance Charges (OBBDITFI)
11. PBDIT/ Interest and Finance Charges (PBDITITFI)
12. Net Cash Accruals/Total Debt (NETTOTA)
13. Total Debt/Tangible Networth (DEBTTNET)
14. Current Ratio (Current Assets/Current Liabilities) – (CURRATIO)