

Mehmet Resul Bilginci

**DECISION SUPPORT SYSTEM FOR RISK MANAGEMENT:
AN APPLICATION IN FINANCE SECTOR**

M.S. Thesis In Industrial Engineering

by

Mehmet Resul BİLGİNCİ

June - 2012

June 2012

**DECISION SUPPORT SYSTEM FOR RISK MANAGEMENT:
AN APPLICATION IN FINANCE SECTOR**

by

Mehmet Resul BİLGİNCİ

A thesis submitted to

the Graduate Institute of Sciences and Engineering

of

Fatih University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Industrial Engineering

June 2012

Istanbul, Turkey

APPROVAL PAGE

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Assist.Prof. Ali Türkyılmaz
Head of Department

This is to certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assist.Prof. Ali Türkyılmaz
Supervisor

Examining Committee Members

Assist. Prof. Ali TÜRKYILMAZ _____

Prof. Dr. Selim ZAİM _____

Assist. Prof. Ömer Fahrettin DEMİREL _____

Assist. Prof. Özgür UYSAL _____

Assoc. Prof. Ali COŞKUN _____

It is approved that this thesis has been written in compliance with the formatting rules laid down by the Graduate Institute of Sciences and Engineering.

Assoc. Prof. Nurullah ARSLAN
Director

June 2012

DECISION SUPPORT SYSTEM FOR RISK MANAGEMENT: AN APPLICATION IN FINANCE SECTOR

Mehmet Resul BİLGİNCİ

M. S. Thesis - Industrial Engineering
June 2012

Supervisor: Assist. Prof. Ali TÜRKYILMAZ

Co-Supervisor: Prof. Dr. Selim ZAIM

ABSTRACT

In this study, Credit Risk Scorecards ,which are used in measuring and managing Credit Risk which is one of the most challenging risks that banks are faced, are mentioned. After having mentioned historical development and model construction phases of credit scorecard models in detail, an example scorecard application ,which is built by using logistic regression method, is illustrated. The data which was used to construct credit scorecard model was obtained from a bank that operates in Turkish Banking Sector. The dataset included credit applications made between March 2008 and November 2010. The applications made between June 2009 and March 2010 were defined as development sample which was used to construct model. Development sample was splitted into 2 parts. 70% of the data was allocated for training and the rest was allocated for testing. GB variable which displays whether credit is good or bad credit in terms of credit risk was used as dependent variable, which also known as target variable, and the other variables were used as independent variables, which also known as predictor variables. The performance of generated model was gauged by statistical metrics such as ROC curve and Gini statistics.

Keywords: Credit Risk, Credit Scorecard Model, Logistic Regression, ROC Curve, Gini Coefficient.

RİSK YÖNETİMİNDE KARAR DESTEK SİSTEMİ: FİNANS SEKTÖRÜNDE BİR UYGULAMA

Mehmet Resul BİLGİNCİ

Yüksek Lisans Tezi – Endüstri Mühendisliği
Haziran 2012

Tez Danışmanı: Yrd.Doç. Dr. Ali TÜRKYILMAZ

Tez Eş Danışmanı : Prof.Dr. Selim ZAİM

ÖZ

Bu çalışmada, bankacılık sektörünün karşı karşıya olduğu en önemli risklerden kredi riskinin ölçülmesi ve yönetilmesi için kullanılmakta olan kredi skorkart modellerine değinilmiştir. Skorkart modellerinin kurulmasında en yaygın olarak kullanılan lojistik regresyon yöntemiyle hazırlanan örnek bir uygulamaya yer verilmiştir. Modelde kurulum aşamalarına detaylı bir şekilde açıklanmıştır. Çalışmanın uygulama bölümünde yer verilen skorkart modelinin kurulumunda Türk Bankacılık sektöründe faaliyet gösteren bir bankanın veritabanından elde edilen Mart 2008 ile Kasım 2010 arasındaki dönemde yapılan kredi başvurularına ait kredinin iyi/kötü olma karakteristiği ve başvuru bilgilerini içeren karakteristikler kullanılmıştır. Elde edilen veriler arasında modelin kurulacağı müşteri portföyünün özelliklerini en iyi yansıtan Haziran 2009 Mart 2010 döneminde yapılan başvurular örneklem seti olarak belirlenmiştir. Belirlenen örneklem seti, katmanlı örnekleme yöntemiyle model geliştirme örnekleme, örneklem setinin %70'i, model test örnekleme, örneklem setinin %30'u, olarak ikiye bölünmüştür. Kredilerin iyi /kötü kredi olma karakteristiği modelde bağımlı değişken olarak, diğer karakteristikler ise bağımlı değişken olarak kullanılmıştır. Kurulan skorkart modelinin performansı ROC eğrisi ve Gini Katsayısı gibi istatistiksel yöntemler yardımıyla test edilmiştir.

Anahtar Kelimeler: Kredi Riski, Kredi Skorkart Modeli, Lojistik Regresyon, ROC Eğrisi, Gini Katsayısı

DEDICATION

To my parents, teachers and friends

ACKNOWLEDGEMENT

I express sincere appreciation to Assist. Prof. Ali TRKYILMAZ and Prof. Dr. Selim ZAIM for their guidance and insight throughout the research.

I express my thanks and appreciation to my family for their understanding, motivation and patience. I would like to thank Seray Gkl, İnci Aksoy and Cansu lgen for their support during my thesis. Lastly, but in no sense the least, I am thankful to all colleagues and friends who made my stay at the university a memorable and valuable experience.

TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZ.....	iv
DEDICATION.....	v
ACKNOWLEDGEMENT.....	vi
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 CREDIT SCORING.....	3
2.1 Descriptions of Credit Scoring	3
2.2 Comparision of Judgemental And Scoring Systems	4
2.3 Historical Development of Credit Scoring	5
2.4 Benefits And Criticism of Credit Scoring	7
CHAPTER 3 SCORECARD DEVELOPMENT METHODS	9
3.1 Decision Trees	9
3.2 Neural Network Models	12
3.2.1 Number of Neurons in the Hidden Layer	13
3.2.2 Applicability of Neural Network in Credit Scoring Models	14
3.3 Logistic Regression	14
3.4 Comparision Of Methods	17
CHAPTER 4 AN APPLICATION OF CREDIT SCORING MODEL	19
4.1 Credit Evaluation Approaches	19
4.2 Defining Predictor Characteristics.....	22
4.3 Defining Dependent Variable	25
4.3.1 Selection Statuses	25
4.3.1.1 Observation Excludes	25

4.3.1.2	Not Taken Up	25
4.3.1.3	Mutual Accept	25
4.3.2	Good / Bad Definition	25
4.3.3	Population Flows	27
4.4	Sample Window.....	28
4.4.1	Sample Window / Observation Period	29
4.5	Model Development	32
4.5.1	Data Partition	32
4.5.2	Initial Characteristic Analysis.....	33
4.5.2.1	Univariate Methodology	34
4.5.2.2	Bivariate Methodology	35
4.5.3	Statistical Measures Used to Characteristic Selection.....	36
4.5.3.1	Information Value.....	37
4.5.3.2	Chi-Square	38
4.5.3.3	Gini Coefficient	39
4.6	Model Building.....	47
4.6.1	Pearson Correlation Analysis	53
4.6.2	Misclassification	56
4.7	Master Scorecard	57
4.7.1	Example Scorecard	58
4.8	Model Performance	60
4.8.1	Gini Coefficient	60
CHAPTER 5 MANAGERIAL IMPLICATION AND CONCLUSION		63
REFERENCES		66

LIST OF TABLES

TABLE

4.1 List Of Characteristics Used In Analysis.....	22
4.2 List Of Characteristics Used In Analysis.....	23
4.3 Good / Bad Definition.....	27
4.4 Structure Of Data File.....	27
4.4 The Dependent Variable Flag Distribution By Application Period.....	31
4.5 Data Partition Node Result	33
4.6 Information Value (F-Statistic) Benchmarks.....	38
4.7 Chi-Square (X^2) Benchmarks	39
4.8 Gini Benchmarks	40
4.9 Interactive Grouping Node Output	42
4.10 Analysis Of Grouped Variables	46
4.11 Regression Model Inputs	49
4.12 Regression Model Output	50
4.13 Correlation Analysis Output	53
4.14 Characteristics And Their Roles	54
4.14 Summary Of Stepwise Selection	55
4.15 Regression Model Elements	55
4.16 Confusion Matrix Regression Output	56
4.17 Logistic Regression Output	57
4.18 Artificial Variable Value For An Applicant	58

4.19 Groups And Woe Of The Model Parameters.....	58
4.20 Score Calculation	59
4.21 Model Performance Output	61
4.22 Roc Benchmark Index	61

LIST OF FIGURES

FIGURE

3.1 Multilayer Feed-Forward Network With One Hidden Layer	13
3.1 Logistic Curve.....	16
4.1 Traditional Credit Underwriting Process	20
4.2 Credit Underwriting Process With Credit Scoring Approach	21
4.3 Sample Window / Observation Period.....	29
4.4 G/B Odds By Application Date	29
4.5 Data Partition	32
4.6 Data Partition Node Properties Window	33
4.7 Interactive Grouping Node	40
4.8 Interactive Grouping Node Properties Window	41
4.9 Regression Node	48
4.10 Regression Node Property Window	49
4.11 Sas Code Node.....	51
4.12 Sas Code Editor Window.....	52
4.13 Correlation Analysis Diagram On Sas Enterprise Guide.....	52
4.14 Regression Node(2)	54
4.15 Confusion Matrix.....	56
4.16 Model Comparison Node.....	60
4.17 Roc For Train And Test Data Samples	61

CHAPTER 1

INTRODUCTION

Banks, which are the most important element of the finance sector, face many risks as the nature of the activities they do. In the economic cycle, banks take on the function of bringing together those are in need and those have fund savings for. While performing this function, the type of risk that banks expose and whose measurement and management is vital for banks in order to be able to survive in today's increasingly competitive financial system is credit risk. Credit risk is defined by Banking Regulation and Supervision Agency (BDDK) as "Credit risk is most simply defined as the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms." (BDDK, 2001a).

Managing risk in a correct manner depends on forecasting the potential risks and taking necessary actions in order to prevent risk from occurring. In traditional approach to evaluating credit applications, banks grant or reject the applications coming from commercial or retail customers in accordance with the credit allocation specialist's opinions. Another method approach used in evaluating loan applications is credit scoring method or more commonly used named Scorecard. Scorecard is used to predict how likely an applicant is to repay the loan he applied for and to answer the question of whether loan will default or not in specific time by classifying the application as good or bad in terms of credit risk based on past statistical experience. Past credit applications are analyzed in order to define the characteristics that have significant effect on discrimination of good and bad applications in terms of credit risk (Anderson, 2007).

The advantages of using scorecard in the process of credit allocation process can be summarized as;

- Better decisions, more quickly and more cheaply
- More consistent credit decisions
- Quantifiable risk management
- Lower transaction costs

The goal of this thesis is to convey an understanding of how credit scorecards are developed and to present a credit scoring application for SME loans. In this study, we used logistic regression technique to build scorecard.

The remainder of the article is structured as follows. In Chapter 2, historical development of credit scoring method is provided. In Chapter 3, the methods such as decision trees, neural networks and logistic regression that can be applied to build credit scoring models are outlined. In Chapter 4, first two different approaches employed in credit underwriting process are summarized. Then, an application of credit scoring model is conducted by using logistic regression with real data which is supplied from a major bank operating in Turkey. In Chapter 5, Implication and Conclusion drawn from the study is expressed.

CHAPTER 2

CREDIT SCORING

2.1 DESCRIPTIONS OF CREDIT SCORING

Credit allocation process, in which credit applications are evaluated, is the one of the most critical and core processes of the banks. This crucial process involves the stages of collecting the credit factors, that are used in assessing credit applications, from the sources such as credit application forms, central bank data, credit bureau data, and the core internal data of bank, analyzing these collected credit factors, and giving final credit decision.

Quality of bank's loan portfolio is closely associated with the competitiveness and profitability of the bank. Quality of bank loan increases as the number of customers having high creditworthiness. Credit scoring is the most important decision support system which is utilized in evaluating the creditworthiness of the applicant. Therefore, it is possible to define the credit scoring as the process of modeling the creditworthiness of applicant (Crook, Edelman and Thomas, 2007).

Furthermore, credit scoring can be defined as transformation of relevant credit factors into numeric measures by using statistical methods to be utilized as a guide in giving credit decision (Anderson, 2007).

Moreover, credit scoring can be described as an analytical model having constructed empirically based on past applications data used to predict creditworthiness of applicant by utilization of probability of default (Malhotra and Malhotra, 2003).

2.2 COMPARISON OF JUDGEMENTAL AND SCORING SYSTEMS

The main idea of credit evaluation is to compare the characteristics of an existing customer, whose loans they have already paid. If a borrower's characteristics are similar to those who failed to fulfill their requirements, the application will generally be declined. If the borrower's characteristics are similar to those who were not defaulted, the application will be accepted. In general, the techniques used to evaluate credit applications can be summarized into two categories. One is judgmental evaluation and the other one is credit scoring (Crook, 1996).

In judgmental evaluation technique, each credit application is assessed by a decision maker who is employed in the bank titled as credit allocation specialist. The performance of credit allocation process in which the judgmental technique is employed depends on experience and the common sense of credit specialist. Since in this method the subjective assessment of credit allocation specialist has a great impact, this method is associated with subjectivity and individual preferences influencing credit decisions. However, taking qualitative characteristics into account is one of major strengths of judgmental evaluation technique.

In credit scoring technique, past credit records are used to derive a qualitative model in order to evaluate the credit applications for the segregation of accept or reject. In credit scoring system, credit evaluation criteria are consistently applied to all credit applications. The scoring system is based on the addition or subtraction of a statistically extracted number of points relating to the applicant's score given to the predictor variables, such as time on a job or the number of credit sources used. And this enables giving credit decisions quickly. Moreover, credit scoring maintain an increase in customer satisfaction since the time required to finalize credit application. On the other hand, credit scoring models have been criticized due to the statistical problems associated with the data to generate the model. However, credit scoring models can be regarded as one of the most successful application of statistics in the field of business (Bailey, 2004).

2.3 HISTORICAL DEVELOPMENT OF CREDIT SCORING

In 1936, in his study, Fisher showed that different plant species could be discriminated by comparing the sepal and petal measured values of the plants. This study is important since it is the first study proving that a group can be classified in statistically significant manner by analyzing the particular common characteristics of the group members (Fischer, 1936).

In 1941, based on the idea of same characteristics could be used in separation of the groups within the same group, Durand did an analysis to find credit factors that have significant influence on classification of credit applications into good and bad classes (Durand, 1941).

In 1930s, some mail-order companies started to use scoring system in order to be able reduce the inconsistencies among the credit analysts. Companies operating in the finance sector encounter difficulties in managing credit risk because of the fact that the conscriptions of credit analysts during the World War II. As a result, there was lack of credit risk experts. Hence, companies had the analysts write down the rules that they used in credit granting decisions. These rules used by non-experts in order to help companies in their credit decision systems which then led to expert systems (Thomas, Edelman and Crook, 2002).

As the World War II ended, consultant firms were founded whose objective was to automate the credit decision systems of the companies who had functions of lending by integrating the statistical models that were developed with the aid of rule sets which were collected from the credit experts before they were conscripted to the military (Wonderlic, 1952).

Increase in the number of applications has increased so much that using the traditional approach in which credit applications are evaluated one by one by credit specialist has become economically impossible. The desire both coming from the ones who apply for the credit and the banks who evaluate the applications to give credit decisions in a reasonable time has encouraged the banks to use credit scoring in the process of evaluating the credit applications (Lewis, 1992). Evaluating the credit card applications via credit scoring approach in a successful manner has led banks to use

credit scoring for home loans and small business loan by the 1980s (Thomas, Edelman and Crook, 2002).

Beaver presented a study in which he made an analysis of potential predictors that could be used to predict the bankruptcy. In his study, he proved statistically that a number of credit factors mostly financial ratios could be used to estimate the probability that a firm goes to bankruptcy (Beaver, 1966).

Altman used discriminant analysis method by using the data of the companies operating in the manufacturing sector in the USA. In his study, he developed a classification model that had an overall accuracy rate to predict the firm failure (Altman, 1968).

Providing a wide range of products by the banks has led to a significant increase in credit applications. This huge increase in demand in the banking sector has resulted in an increase in credit scoring applications. Banks have begun to evaluate almost every credit applications for consumer loans, credit card, small business loans and home loans (Emel, et al. , 2003).

Credit scoring models are built by using the historic data of the accepted credit applications in the past. For this reason, the credit scoring model cannot learn the properties of the rejected applications. Since in real life, there are applications which will be rejected, the accuracy of the credit scoring model decreases. Feelders, in order to solve this problem used both accepted and rejected applications to develop a mixture model. In his study, in which he deployed reject inference method, he showed that performance of the model increases in a remarkable way by the inclusion of the rejected applications (Feelders, 2000).

Banks are getting more concentrated to carry a variety of empirical studies in order to be able to benefit the classification techniques, which they have already implemented successfully in distinguishing good and bad credit applications, to develop strategies that maximize the profit gaining from their both new and existing customers (Finlay, 2009).

2.4 BENEFITS AND CRITICISM OF CREDIT SCORING

Since credit scoring models include only the credit factors which have significant effect in discriminating the good and bad applications, credit scoring models need less information to evaluate the credit applications. However, judgmental evaluation techniques are relied on subjective assessment of credit analyst whose evaluation has no statistical analysis. Rejected applications are also analyzed to find out how they would have performed if they had been granted in credit scoring models. Judgmental methods are based on only the features of the accepted applications.

In credit scoring models, the decision of granting or rejecting the application is given by estimating the probability of the application being bad or defaulted. On the other hand, judgmental methods are biased towards awareness of bad payers. Credit scoring models are built on large samples that credit allocation specialist can not keep in mind. In credit scoring technique, the association between the credit factor and the repayment behavior of the applicant is displayed whereas this association can not be demonstrated in judgmental evolution technique. In credit scoring approach, a large number of characteristics can be involved in the model however to evaluate a combination of several characteristics of the applicant at the same time is impossible for a credit analyst where the judgmental evaluation method is employed. Another benefit of credit scoring models, the data used to generate the model can be analyzed easily and a new model whose output are almost the same with the previous one can be built by another credit scoring models specialist. On the other hand, it is very unlikely to give the same credit decision for the same credit application by different credit allocation specialist.

Some other advantages of credit scoring model can be summarized as follow: more effective processing time, support for decision making process, minimization of the cost association with credit evaluation process, fewer errors, measuring and managing credit risk more objectively and correctly with using variables whose applicability has been proved through statistical analysis, less information needed to evaluate credit applications, enabling to adjust the credit approval rate in accordance with bank's risk appetite and external factors affecting the banking sector.

One of the criticism against the credit scoring model is to involve the characteristics that have been not fully justified that they have a significant effect on creditworthiness of an applicant. Also, when the credit scoring models are generated, sometimes economic factors are ignored to add into the scoring model. In addition, due to the misclassification problem of credit scoring models can sometimes flag the applicant as bad who is actually good applicant. Furthermore, the credit scoring models are not standard, differ from market to market, are expensive to supply from the consulting firms which are expertise in building credit scoring models (Al Amari,2002).

Credit decision is given in an environment where the credit scoring approach is employed as follows: a customer applies for the loan with some information that are important for bank since it converts the customer's application information into credit factors. Customer's application is evaluated in scoring system by simply summing the points received on the various features to have a total score. This score can be treated in different ways depending on the credit policies. Generally, applicant's total score is compared with to a predefined cut-off score. If the customer' total score is greater than the cut-off level, then the application is approved otherwise the application is rejected. Sometimes, banks design their credit scoring system based on a two stage process. If the customer's total score is greater than higher cut-off, credit application is automatically approved, if it is smaller than the lower cut-off, application is automatically declined. If the score is between two cut-off points, the application is evaluated by credit allocation specialist.

Credit scoring models are only as good as the original features, general limitation of these models is that the historical data is used. Although it is assumed that model variables or their weights or both of them to remain constant, unless model is updated in particular frequencies, accuracy of the model diminishes. This issue can be overcome if the banks monitor the Type I and Type II errors regularly and apply new or updated model to remain the model adequacy at the original specification level (Heffernan, 2005).

CHAPTER 3

SCORECARD DEVELOPMENT METHODS

3.1 DECISION TREES

In this method, which was initially developed by Breiman for classification problems, the predictor variables in the application data are splitted into subsets, and each subset is assigned as Good or Bad accordance with the target variable by looking at the majority in the node (Breiman et al., 1984).

A decision tree can be defined as a flow chart where each internal node represents a test on an attribute, each branch denotes an outcome of the test, and each terminal node displays a class label (Han and Kamber, 2006).

A decision tree procedure can be applied into credit scoring problems as in the following structure. Application dataset, let's call it Application Data A, is splitted into two subsets by taking the previous samples into account. These two new subsets shows a more homogeneous distribution than the original set in terms of credit risk. Then, in order to be able to create more homogeneous subsets, these new subsets are splitted into new subsets, and the process is repeated. Hence, this method is also called as recursive portioning. The process of creating new subsets is continued till the subsets meet the requirement for being terminal node. When the requirements are met, each terminal node is assigned A_G which denotes class of Good Accounts and A_B which denotes class of Bad Accounts (Safavian and Landgrebe, 1991).

In classification trees, there must be given three decisions:

- rules to use split the sets into two - the splitting rule,
- how to determine whether the set is terminal node or not - the stopping rule,
- define how to terminal nodes are assigned into good or bad categories

Assignment of terminal nodes into categories is relatively easy decision to make. If the majority of the node is good, then node is defined as good category on the other hand if the majority is bad observations in the node, then the node is defined as bad category. Another way to flag the node good or bad is to minimize the cost of misclassification. Let D be the cost of misclassification of a bad account as good and let L be the cost of misclassification of a good account as bad account. In circumstances where ratio of number of goods to number of bads exceeds the D to L , node is classed as good thus the misclassification cost is minimized (Thomas, Edelman and Crook, 2002).

In order to determine the splitting rules that are used to create subsets, all possible splits are examined and the split rule which maximizes the classification between two groups is chosen. The most widely used splitting rules in classification problems are Information Gain, Gain Ratio and Gini Index. These splitting rules determine the best variable that is used to split the set by measuring the impurity information of the node.

Information gain is the most popular measure originated from information theory. Input variable with the highest information gain is selected in order to decrease the information need of descendent nodes and represents their impurity. Let $I(S)$ be the impurity of a node S and $n(S)$ be the number of observations in node S . The information impurity (or entropy impurity) will be defined as follows:

$$I(S) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (3.1)$$

where p_i is the probability of class i in node S and is estimated by $n(S_i)/n(S)$. $I(S)$ gives the average information value needed to classify the observations in node S , which is based on the proportions of observations of each class. The expected information to find an exact classification after further partitioning of node S , is measured as follows:

$$E(A) = \sum_{j=1}^v \frac{n(S_j)}{n(S)} \times I(S_j) \quad (3.2)$$

where j is the class label of new feature A until v , and S_j is the subsequent node of S . The smaller the expected information, the greater the purity of partitions.

The difference between the original information impurity and expected information is the information gain:

$$Gain(A) = I(S) - E(A) \quad (3.3)$$

Knowing the value of feature A will reduce the information requirement and the attribute with the highest information gain will be chosen for splitting.

Information gain has a drawback that it prefers to select features with large number of values. Because a feature with large number of values will provide more partitions and will reduce the impurity more than another one with small number of values will. For instance, a unique identifier will provide expected information of 0 and therefore the information gain will be maximal (Han and Kamber, 2006).

To overcome this bias, Gain Ratio is introduced, which applies a kind of normalization to information gain using split information. The split information is defined as follows:

$$Split(A) = \sum_{j=1}^v \frac{n(S_j)}{n(S)} \times \log_2 \left[\frac{n(S_j)}{n(S)} \right] \quad (3.4)$$

For each outcome, the number of observations having that outcome with respect to the total number of observations in S is considered. On the other hand, information gain measures the information with respect to classification that is obtained based on the same partitioning. The maximum value of gain ratio is selected for splitting which can be calculated with the following equation:

$$GainRatio(A) = \frac{Gain(A)}{Split(A)} \quad (3.5)$$

Another measure of impurity is the Gini index that considers a binary split for each feature. For discrete features the subset, and for continuous features the split point that gives the minimum Gini index is selected for splitting. Gini index is expressed as follows:

$$Gini(S) = 1 - \sum_{i=1}^m p_i^2 \quad (3.6)$$

While creating decision trees, another decision point that must be determined is stopping rule. In some decision tree applications, decision tree is continued to grow till each leaf node has a particular number of observations. The minimum number of case that must be in the node is somewhere between five and fifty for credit scoring models. However, even the condition of this minimum size of the leaf node results in over-fitting problems. To tackle this obstacle, pruning is applied once the decision tree is built. Pruning is the transaction by which some of leaf nodes are removed from the decision tree in order to be able to decrease the size of decision tree.

The best way to apply pruning is, splitting the development sample into training and validation samples. By using the training sample, a decision tree with a maximum possible size is built, and by using the validation sample the performance of the generated model is assessed. Then, a leaf node is removed from the decision tree and the performance of the model is gauged again. If the removal of the leaf node leads an increase in the model performance, then leaf node is not added back into the tree but if removal of the leaf node leads a decrease in the model performance, then leaf node is added back into the tree. Removal of the leaf nodes continue till the point where no performance increase is observed (Thomas, Edelman and Crook, 2002).

3.2 NEURAL NETWORK MODELS

Neural networks are mathematical models that simulate the biological nervous system. The common characteristics of neural networks and biological neurons are parallel processing of information and learning and generalizing from experience (Patterson, 1996).

Neural networks, which came out from studies for artificial intelligence, began to gain importance in the 1980s (Rumelhart, Hinton and Williams, 1986).

Neural networks can be used when there is a poor relationship between inputs and outputs and they are also successful on incomplete and noisy data classifying patterns on which they have not been trained (Han and Kamber, 2006). The main feature of neural networks is their ability to represent non-linear features of the data through the inclusion of two or more hidden nodes contained within the hidden layer of the network (Finlay, 2010).

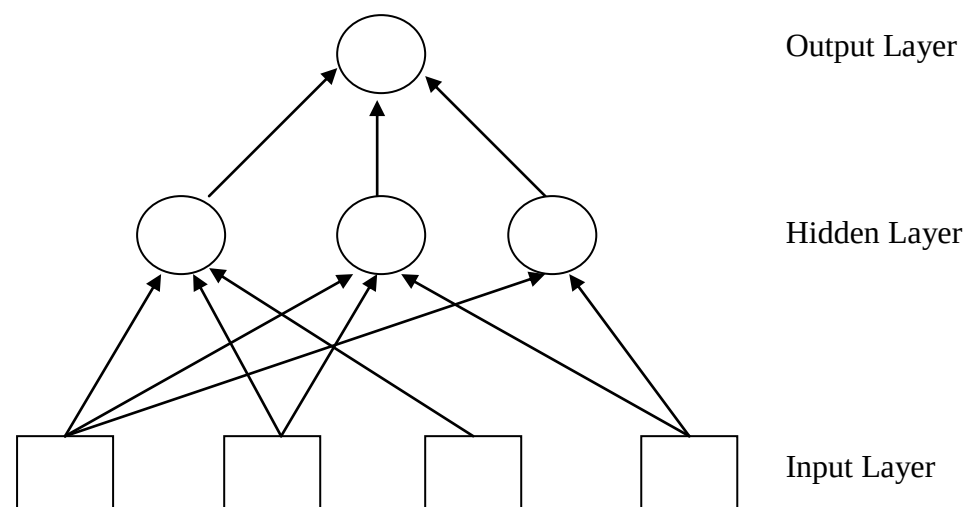


Figure 3.1 Multilayer Feed-forward Network with One Hidden Layer.

Figure 3.1 represents a multilayer feed-forward network with one hidden layer. Feed-forward networks include nodes that are connected only to the nodes in the next layer. None of weights returns back to an input or output node of a previous layer. However, in recurrent networks, nodes in the same layer can be connected to each other.

3.2.1 Number of Neurons in the Hidden Layer

In neural networks, the optimum number of neurons in the hidden layer differs from problem to problem. In order to be able to determine the best number that is used in the hidden layer, there must be generated so many experiments. As the number of neurons in the hidden layer increases, model's ability to capture the non-linear association also increases. On the other hand, the number of parameters that model

needs to make estimation increases as well. This increases the time to train the network and increases the probability of over-fitting occurring.

When neural network method is used for building the credit scoring model, since the discriminatory power of the linear model is good and sometimes even better than the neural networks there is no need to use so many neurons in the hidden layer (Finlay, 2010).

3.2.2 Applicability of Neural Network in Credit Scoring Models

A huge amount of academic research and lots of publicity, about how these neural networks can solve complex problems and predict different types of behavior with a high degree of accuracy. In some problem domains the promise of improved prediction has been well founded and neural networks have been demonstrated to deliver marked benefits in comparison to many other forms of model. However, when it comes to using within financial services, the evidence is less supportive. There is some evidence that neural networks can make better predictions than linear models and decision trees by a small, albeit statistically significant, margin (West et al., 2000).

However, many empirical studies covering direct marketing, credit scoring and insurance risk have found the performance of neural networks to be about the same, or even marginally worse, than linear models (Linder et al., 2004).

Compared to linear models and decision trees, artificial neural networks are complex. This leads to greater costs in terms of the effort required to develop and implement them within operational environments. Understanding how an artificial neural network model makes its decisions is also difficult, due to the interdependent nature of the neurons and weights within the network. This means they tend not to be used for decision making in areas where explicable models are required.

3.3 LOGISTIC REGRESSION

Logistic regression is a predictive modeling technique, a special form of regression, which is developed for binary categorical targets. It finds out the

relationship between two or more independent variables and a single binary dependent variable.

Logistic regression differs from multiple regression by the type of dependent variable, estimation methods, and assumptions about the underlying distribution. It violates two assumptions of linear regression: Instead of a normal distribution, the error term of discrete variables follows the binomial distribution and the variance of a binary variable is not constant (Hair et al., 1998).

Logistic regression predicts the probability of an event occurring, however to hold the predicted value in a range of zero and one, it uses an S-shaped logistic curve that symbolizes a relationship between independent and dependent variables.

Figure 3.1 illustrates a logistic curve drawn for a single independent variable x . At very low values of the independent variable, probability approaches to zero. For a certain range of intermediate values, probability increases rapidly and at very high values, it approaches to one, but never exceeds these boundaries.

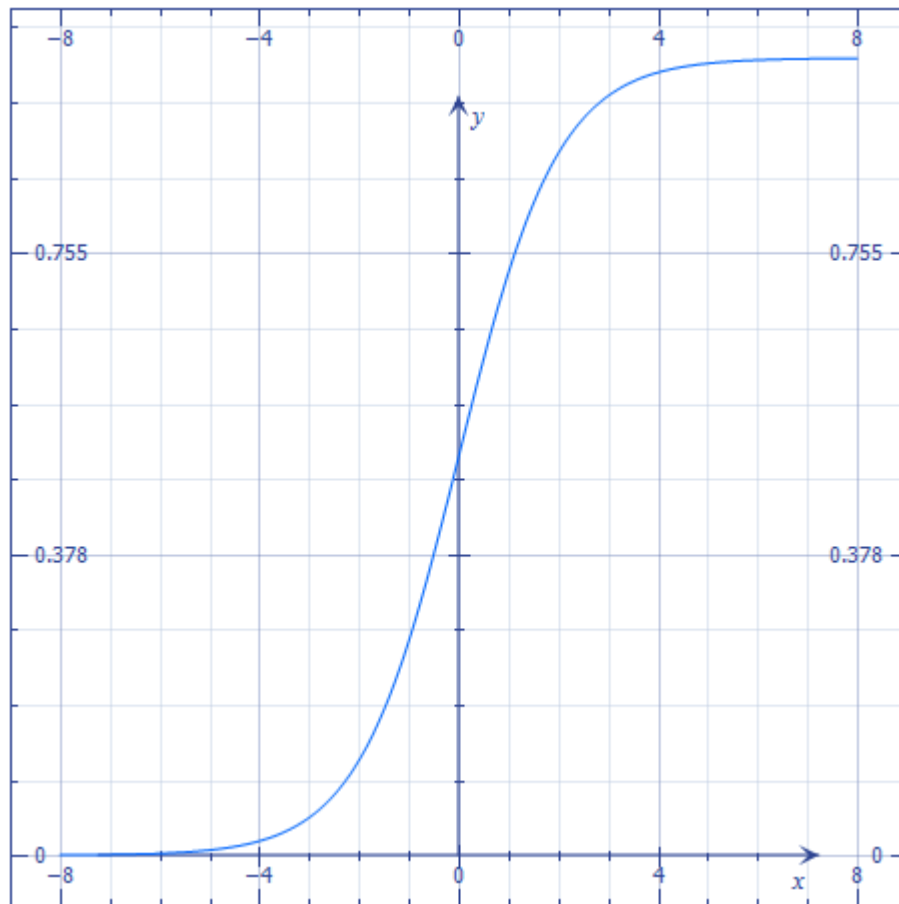


Figure 3.1 Logistic curve.

Given several predictor variables, logistic function is defined as the following equation:

$$p_i = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (3.7)$$

where :

α is the intercept

β_j is the coefficient of independent variable x_j

p_i is the probability of the binary dependent variable y being 1.

In order to obtain the unknown parameters, the intercept and the coefficients, logistic regression uses maximum likelihood method because of its nonlinear structure. Therefore, likelihood value is used to calculate the measure of overall model fit. On the

other side, multiple regression mostly uses least squares method to estimate the coefficients and sum of squares for the measures of overall model fit.

To transform the values of binary dependent variable into logistic curve that represents the probability of an event, logit transformation is used.

$$\text{logit}(p_i) = \log\left[\frac{p_i}{1-p_i}\right] = \alpha + \beta_1\chi_1 + \beta_2\chi_2 + \dots + \beta_n\chi_n \quad (3.8)$$

Logit transformation is the log of odds ratio, which is an important measure for logistic regression that compares the probability of an event with the probability of an event not happening, and is calculated as follows:

$$\text{Odds ratio (OR)} = \frac{p}{1-p} \quad (3.9)$$

The estimated coefficient β_j in logit function represent the changes in log odds ratio for a unit change in χ_j . Positive coefficient values increase; negative coefficients decrease the predicted probability. For instance, if β_j is positive, its transformation will be greater than 1, the odds ratio will increase and therefore, the model will predict $y = 1$ better.

3.4 COMPARISION OF METHODS

Which model will have best performance depends on details of the problems such as structure of data, characteristics used in modeling, boundary to which discrimination can be done by using those characteristics, and the aim of classification. In circumstances where the classed are not defined exactly, then the model used to discriminate the goods and bads will have flat function, so that discriminative power of the model will be accurate. In these cases, using neural network model results in over fitting problem.

Accuracy of the classification is only one dimension of the performance. The speed of classification process, the speed of model updating, the understandability of classification methodology, and why it has its conclusion are the other aspects of the

performance. For an applicant, banks that have instant credit decision systems are more appealing rather than the ones with credit decision systems that have take several times to give a credit decision. For a scoring model being sensitive to a drift in population, when it decreases, an ability to update the model in a fast and cheaply manner is also important. When it comes to these situations, logistic regression or decision tree is more attractive. These models also enable to explain why model has reached its credit decision more easily (Hand, D.J. and Henley, W.E., 1997).

Apart from the studies claim that improvement of model performance is close related with method used to construct the model, there are also studies claim that the significant improvements can be achieved by the addition of new and more predictive characteristics (Henley, W.E and Hand, D.J., 1996).

CHAPTER 4

AN APPLICATION OF CREDIT SCORING MODEL

4.1 CREDIT EVALUATION APPROACHES

Credit Underwriting Process in which loan application of the customer triggers the process and ends whether approve or reject the application is one of the core processes in banks. Credit Underwriting Processes can be grouped into two categories in terms of methods used in giving credit decisions one is Traditional Credit Evaluation method and the other one is Automated Credit Evaluation (Crook, 1996).

In Figure 4.1, credit underwriting process in which the traditional method used is illustrated. In this approach, credit application of the customer is assessed preliminary by branch and the applications which are pre-approved by the branch are directed to headquarter for the final decision. The pre-approved applications are evaluated and the final credit decision is given by the credit underwriting specialists in headquarter.

In today's banking sector where the credit applications are getting higher and higher substantially, it is quite difficult and even impossible to compete with the rivals by employing this traditional approach to manage the credit underwriting process due to the following reasons. First of all, hence the credit decision is made by relying on the subjective judgments, the applications, which have similar properties sometimes even the exact properties, are given different decisions. That causes a significant increase in inconsistency of credit evaluation decisions. Moreover, because of the fact that credit applications are evaluated manually, the time spent for decision making is too much which has potential to result in customer dissatisfaction and ultimately ends with loss of market.

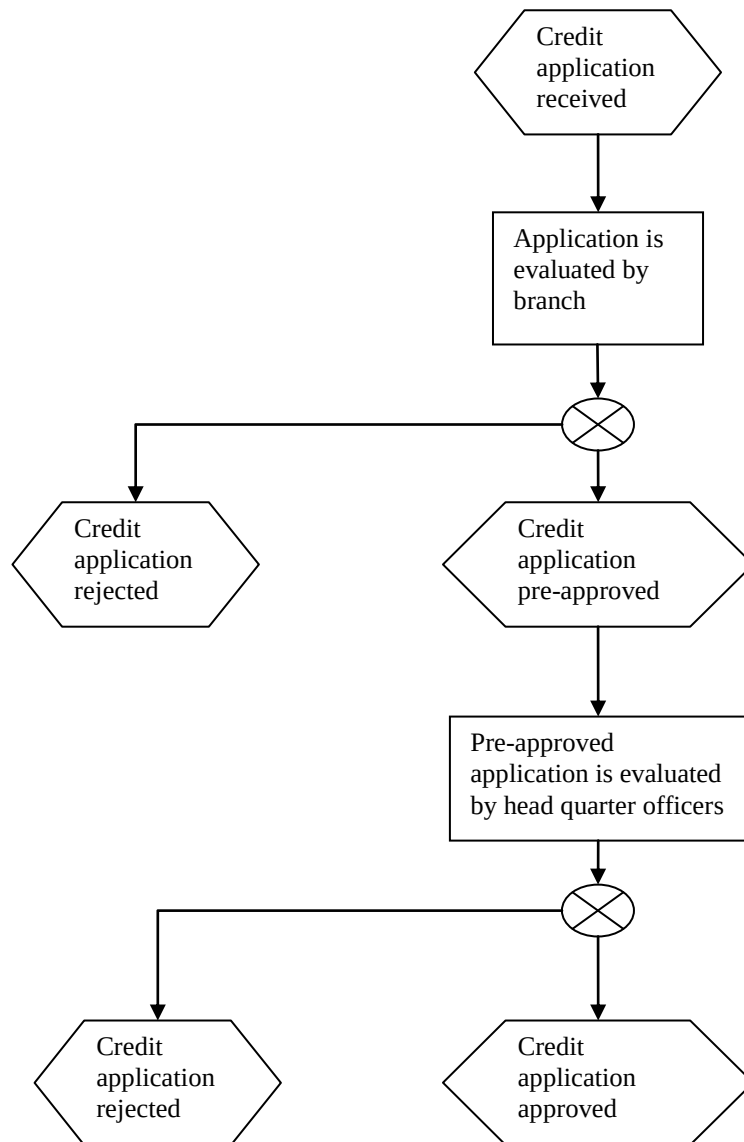


Figure 4.1 Traditional Credit Underwriting Process.

Credit Underwriting Process, in which credit scoring approach, whose applicability in the credit evaluation is proved statistically in this paper, is gained, is illustrated in the Figure 4.2.

Unlike to the traditional approach, in this approach, when the credit application is received, the credit scoring model (scorecard) is run and one of the decisions of approve, reject or review manually is given. As a result, each application is

treated equally and with objective credit factors which enables to increase the consistency of credit decisions and reduce the time spent for credit evaluation. Since, with a scorecard which has a high predictive power increases the automatic credit decision making, the cost of credit underwriting process decreases significantly.

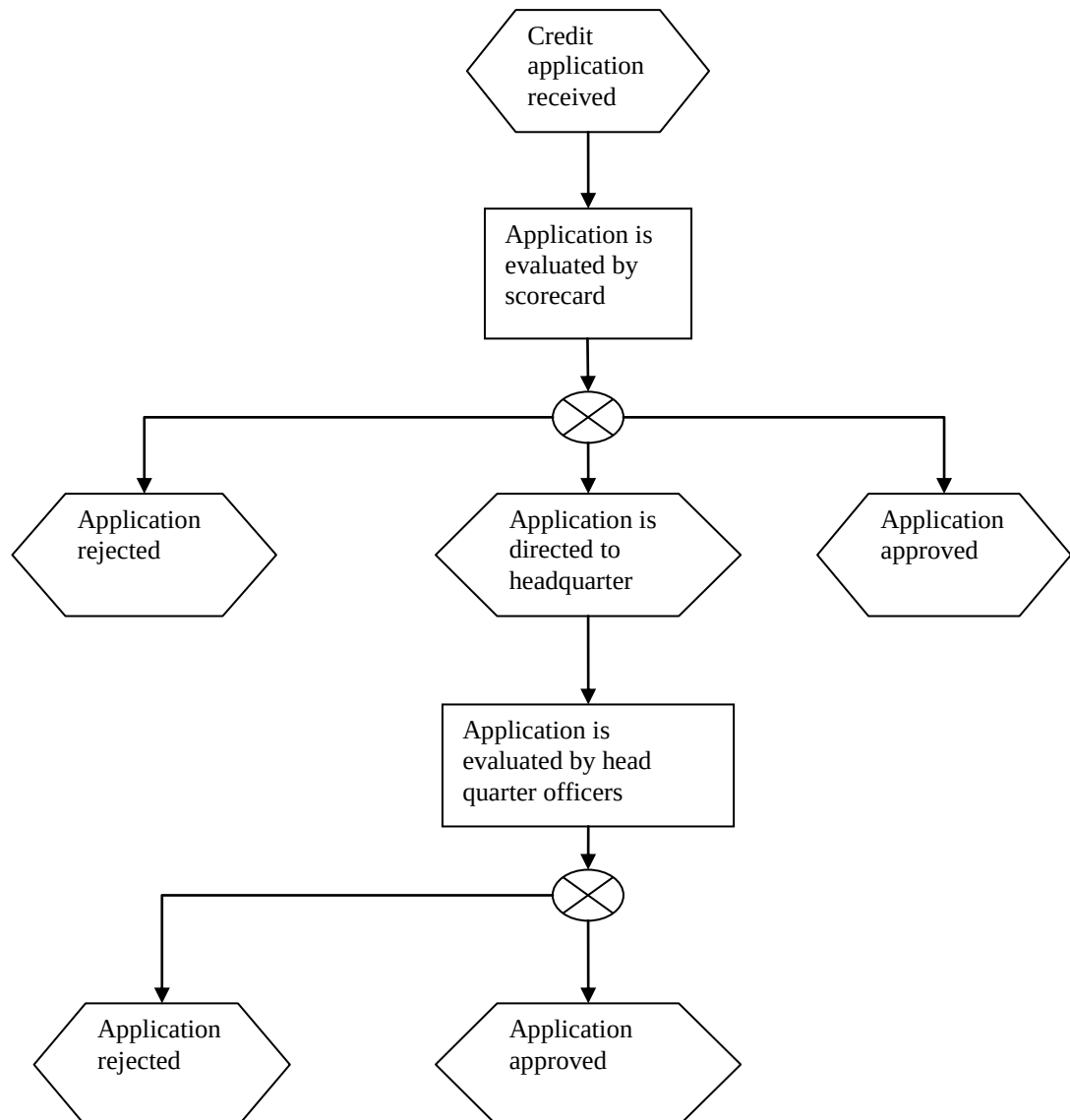


Figure 4.2 Credit Underwriting Process with Credit Scoring Approach.

4.2 DEFINING PREDICTOR CHARACTERISTICS

This stage of the scorecard development process in which the potential credit factors that have significant in discriminating of good and bad credits are investigated is also known as preparation of data dictionary. Because of the fact that data dictionary preparation stage is the first practical step in order to build scorecard model, it is a key stage as any mistakes made in this step may result in situations that is very difficult or sometimes even impossible to compensate (Anderson, 2007).

The first empirical study based on statistical methods whose objective is to build statistically significant credit scoring model was made by Durand. In the initial stage of his study, credit specialists of the banks that participated to the study were required to write down the credit factors they consider important while giving credit decisions and to put them in order of importance. The obtained results were categorized into 5 main groups as shown Table 4.1 (Durand, 1941).

Table 4.1 List of Characteristics Used in Analysis.

Category	Credit Factor	Index
Vocational Characteristics		100
	Work performed	25
	Industry and employer	4
	Stability	71
Financial Characteristics		103
	Assets	21
	Liabilities	46
	Income balance	36
Moral Characteristics		162
	Pat payment record	63
	Character and reputation	59
	Credit rating	40
Personel Characteristics		30
	Age	5
	Marital status	4
	Number of dependents	18
	Stability of residence	3
Loan Characteristics		27
	Security	13
	Duration	14

The credit factors which defined with Durand's study constitute the source of elements that in today's banking sector.

The data sources, in which potential credit factors that are used to estimate whether the credit will be good or bad are listed, can be categorized under the following headings:

- Application forms
- Credit bureau
- Internal systems
- Data obtained from the unit of the central bank of the country

Data used in application, which had both approved and rejected applicants' application data as well as the payment performance data for approved applicants is ,supplied by a major bank that operates in Turkey.

The application data fields supplied by bank are;

Table 4.2 List of Characteristics Used in Analysis.

No	Field Name	Field Info
1	process_no	Process Number
2	appl_month	Month of application
3	appl_date	Application date
4	birth_date	Birth date of main shareholder
5	Age	Age of main shareholder
6	appl_stat	Application status
7	existing_cus	Indicator to define whether main shareholder is an existing customer or not.
8	cus_tenure	Number of years that main shareholder is existing customer
9	gender	Gender of main shareholder
10	Marital_stat	Marital status of main shareholder
11	spouse_work	Indicator to define whether main shareholder's spouse is working or not.
12	firm_op_own	Ownership of operation facilities
13	firm_city_code	City code of company location
14	drwn_cheque	Indicator to define whether there is drawn cheque record of company partners
15	firm_web	Indicator to define whether company has a web page or not.

No	Field Name	Field Info
16	edu_stat	Education status of main shareholder
17	firm_age	Age of company
18	sector	Sector that the company operates
19	avg_term	Collection period of accounts receivable
20	dom_sales_con	Domestic sales condition
21	msh_lien_rec	Indicator to define whether main shareholder has a lien/executive record
22	exp_y	Experience of main shareholder in years
23	exp_m	Experience of main shareholder in months
24	Exp_ym	Experience of main shareholder in years and months
25	op_dur_y	Operation duration at the previous location in years
26	op_dur_m	Operation duration at the previous location in months
27	o op_dur_ym	Operation duration at the previous location in years and months
28	Dependant_num	The number of dependants of main shareholder
29	real_est_rec	Indicator to define whether main shareholder has real estate
30	res_stat	Residential status of main shareholder
31	residence_y	Time at current residence in years
32	residence_m	Time at current residence in months
33	residence_ym	Time at current residence in years and months
34	open_credit_rec	Indicator to define whether applicant has an open credit in the bank
35	supplier_risk	Indicator to define whether applicant has one supplier risk
36	col_prvd	Indicator to define whether collateral is provided or not
37	bus_card_req	Indicator to define whether applicant request business card
38	num_empl	Number of employees working in the company
39	cus_type	Customer type of applicant
40	id_type	Identity type of main shareholder
41	msh_oth_inc	Indicator to define whether main shareholder has another income
42	nationality	Nationality of main shareholder
43	msh_cc_limit	Credit card limit of main shareholder
44	msh_oto	Indicator to define whether main shareholder has own car
45	p_rej	Policy reject code
46	payment_type	Payment type that applicant has chosen for the applied credit
47	cc_stat	Credit card status of main shareholder
48	msh_member_org	Number of organization that main shareholder is a member
49	mgmt_risk	Indicator to define whether there is management risk in the company
50	GB	Indicator to define whether applied credit is bad or good

4.3 DEFINING DEPENDENT VARIABLE

4.3.1 Selection Statuses

The development sample terms associated with sample size and development sample will be discussed in detail in the later sections on the paper, may consist of the different statuses which are described as follows.

4.3.1.1 *Observation Excludes*

Credit records belonging to Staff loans, VIPs loans, pre-approved loans, etc., whose application score will have no impact on decision making must be removed from the development sample in order to prevent predictive power of the scorecard adversely affected.

4.3.1.2 *Not Taken Up*

Sometimes the approved applications are not activated because of some particular reasons such as applicant does not like the terms, applicant takes a better offer from another bank or applicant has no need for the loan anymore. Credit records that fall into this group must be removed from the development sample.

4.3.1.3 *Mutual Accept*

It is the group in which there are credit records that were approved by the bank and activated by the applicant. Performance of the credit accounts in this group is used as dependent or target variable in the stage of scorecard modeling.

4.3.2 Good / Bad Definition

Credit scoring models are developed with the assumption that past performance will have an influence on the future performance. Based on this assumption, performance of the accounts that were opened in the past is analyzed and try to estimate the performance of the future accounts.

The credit accounts that were analyzed can be categorized into 3 groups named good, bad and indeterminate in terms of their performance.

The following considerations are taken into consideration during the definition of bad (Siddiqi, 2006).

1. Definition of the bad loans should be parallel with the objectives of the bank. If the target is to increase profit, the definition of bad should be at delinquency point where the bank gains no profit.
2. In order to make the distinction between good and bad credit more obvious, it may be specified a 120 delinquent days for a credit account to be counted as bad. However, such a definition since could detract the number of bad loans from the development sample may result in model becomes over fitted.
3. In the case of a 30 delinquent days required for a credit account to be counted as bad, it becomes difficult and difficult to distinguish between good credits and bad credits. That results in a decrease in the predictive power of the credit scoring model.
4. Definition of bad must be compatible with the requirements of the regulatory institutions such as Bank for International Settlements. Based on the new Basel II Capital Accord, the definition of bad or default is as follows:

“A default is considered to have occurred with regard to a particular obligor when one or

More of the following events have taken place.

- (a) It is determined that the obligor is unlikely to pay its debt obligations (principal, interest, or fees) in full;
- (b) A credit loss event associated with any obligation of the obligor, such as a charge-off, specific provision, or distressed restructuring involving the forgiveness or postponement of principal, interest, or fees;
- (c) The obligor is past due more than 90 days on any credit obligation; or
- (d) The obligor has filed for bankruptcy or similar protection from creditors.”

Credit accounts with no delinquency are marked as Good, the ones have between 30 and 60 delinquency days are marked as Indeterminate. Credit scoring model is built by only using credit accounts marked as Good and Bad.

The definition of good and bad clients is the first step of risk scoring model. Besides good and bad clients, indeterminate clients are also determined to separate good and bad clients robustly from each other.

It was decided to use the definition below for GB Flag definition.

Table 4.3 Good / Bad Definition.

G/B Indicator	Good/Bad Indicator Description	Good/Bad Flag
1	Written off	Bad
1	Legal/Default	Bad
1	Delinquency days ≥ 90	Bad
0	Delinquency days = 0 and no Written off or Legal/Default record	Good
8	Inactive Accounts	Indeterminate
0	Closed Good	Good
9	Reject	Rejects

4.3.3 Population Flows

The structure of all data and sample window is shown below;

Table 4.4 Structure of Data File.

Portfolio	Total		Sample Window	
	Frequency	Percent	Frequency	Percent
Good	54440	18%	13209	21%
Bad	24262	8%	3683	6%
Indeterminate	64599	22%	16772	27%
Reject	156699	52%	28040	45%
Total	300000		61704	

4.4 SAMPLE WINDOW

The data set which is used in model building should reflect the properties of the credit applications when the credit scoring model is put into service is required to be as current as possible. Number of the records can be huge on the portfolio of which the credit scoring model is constructed. And these records are saved on the tables which are located on multiple IT systems. Work is required to fetch related data from each source and join it to create a table of data that can be used for statistical analysis and model building. However, accuracy of this work decreases as the recording date of the data goes older. Due to all these concerns, instead of using all the data available of the portfolio for which the credit scoring model is established, a sample that represents the population best is defined and credit scoring model is constructed using this sample (Finlay, 2010).

Using a sample which consists of around 1500-2000 cases of goods, bads and rejects where the reject inference is used is generally enough and used widely to develop credit scorecard (Siddiqi, 2006).

The application data between March 2008 and November 2010 was used. Outcome period is typically a 12 to 24 months to be sufficiently long to allow the accounts to mature and bads to emerge. Additionally, it should be up to date (recent) so that the data is representative of the current population.

A sample window defines the time period over which the applications included in the analysis were received. The purpose of using a sample window rather than the whole data is to include only those applications that were made sufficiently long time ago so that they are mature enough to return stable analysis results.

The sample window is chosen according to the below criteria:

1. Using data from a period as recent as possible to avoid portfolio changes;
2. Including accounts with mature performance only
3. 12 months of sample period to take into account all seasonality effects.

As a result of the performance maturity analysis explained above, the sample window on which the scoring model developed is chosen to be between June 2009 and March 2010 (10 months).

4.4.1 Sample Window / Observation Period

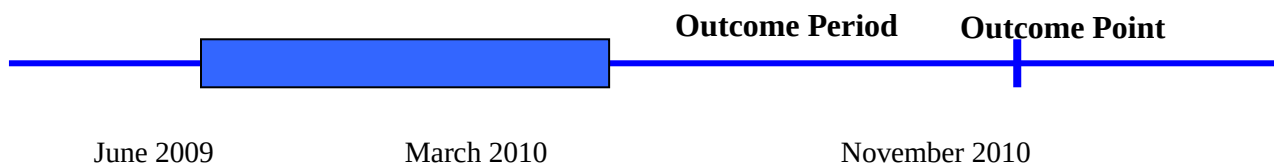


Figure 4.3 Sample Window / Observation Period.

Scorecard which is developed on these mature applications covered in the sample window is then tested on Recent Sample in order to observe whether the score distribution generated by the scorecard is similar to the score distribution in Sample Window. Recent Sample for this particular data set is chosen to be between April 2010 and November 2010.

The G/B Odds that are used in the graph below are calculated from the performance of the applicants in given periods. The application date of the applicants takes place on the X-axis while G/B Odds is on the Y-axis;

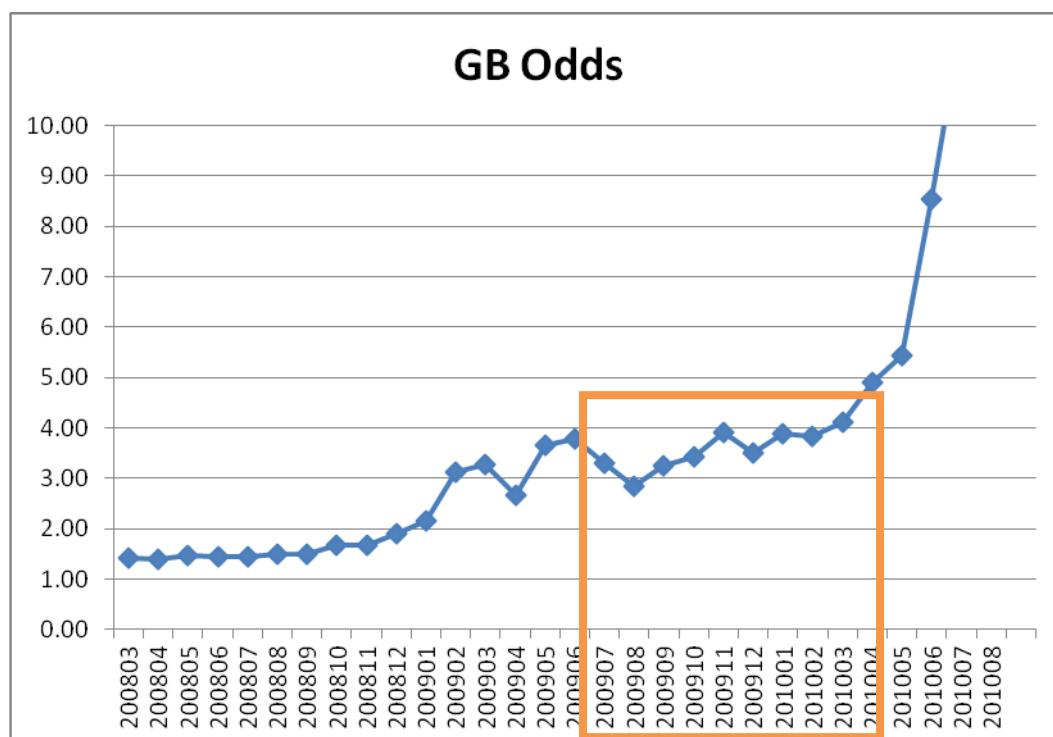


Figure 4.4 G/B Odds by Application Date.

As it is seen in the graph above, the last month that the G/B Odds are stable (in other words applications are mature) is March 2010, after this month, the G/B Odds has

an increasing trend. Therefore this interval (shown in orange frame) is chosen as Sample Window.

The sample window data is randomly broken into development sample (80% of sample window) and test sample (remaining 20% of sample window). Records in the development sample are used in generating the model and records which fall into test sample are used to obtain a final, unbiased estimate of the generalization error of the model.

The Good/Bad Flag distribution among application months is given in detail below:

Table 4.4 The Dependent Variable Flag Distribution by Application Period.

Application Period	Good	Bad	Indeterminate	Reject	Total	GB Odds	Reject Rate	
200803	4175	2913	3490	10243	20821	1.43	49%	
200804	4417	3179	4830	12128	24554	1.39	49%	
200805	5003	3381	7724	13068	29176	1.48	45%	
200806	2849	1985	5517	8840	19191	1.44	46%	
200807	1489	1028	1542	4334	8393	1.45	52%	
200808	1474	986	1242	5104	8806	1.49	58%	
200809	2196	1474	1669	7747	13086	1.49	59%	
200810	1977	1175	1626	7208	11986	1.68	60%	
200811	1336	793	891	4404	7424	1.68	59%	
200812	1263	662	902	6770	9597	1.91	71%	
200901	1441	670	820	11255	14186	2.15	79%	
200902	1388	443	1096	6425	9352	3.13	69%	
200903	1297	396	1206	3447	6346	3.28	54%	
200904	968	362	707	2766	4803	2.67	58%	
200905	901	246	665	1920	3732	3.66	51%	
200906	1030	272	822	1893	4017	3.79	47%	Sample Window
200907	926	281	830	2009	4046	3.30	50%	
200908	990	347	1274	1981	4592	2.85	43%	
200909	1283	396	1540	2388	5607	3.24	43%	
200910	1875	546	3190	4144	9755	3.43	42%	
200911	1224	312	1601	2518	5655	3.92	45%	
200912	1347	384	1670	3206	6607	3.51	49%	
201001	1167	301	1382	2594	5444	3.88	48%	
201002	1508	393	2110	3477	7488	3.84	46%	
201003	1859	451	2353	3830	8493	4.12	45%	
201004	1367	279	1714	3004	6364	4.90	47%	
201005	1295	238	1690	2914	6137	5.44	47%	
201006	1700	199	2677	3819	8395	8.54	45%	
201007	1196	100	1485	2737	5518	11.96	50%	
201008	1099	58	1596	2881	5634	18.95	51%	
201009	1092	12	1762	2904	5770	91.00	50%	
201010	814	0	1647	2459	4920	0	50%	
201011	494	0	1329	2282	4105	0	56%	

4.5 MODEL DEVELOPMENT

4.5.1 Data Partition

There are various ways to split the development (sample on which the scorecard is developed) and validation (“hold”) datasets. Normally, 70% to 80% of the sample is used to develop each scorecard; the remaining 20% to 30% is set aside and then used to independently test or validate the scorecard. Where sample sizes are small, the scorecard can be developed using 100% of the sample and validated using several randomly selected samples of 50% to 80% each.

In this application, development sample was splitted into two parts by using Data Partition node of SAS Enterprise Miner. Stratified partition method was used. Using stratified partitioning, variables are specified to form subgroups of the total population. Within each subgroup, all observations have an equal probability of being written to one of the partitioned data sets. Stratified partition is performed to preserve the subgroup proportions of the population within each partition data set. This improves the classification precision of fitted models.

Training data was set to 70% to build model and test data was set to 30% to evaluate how well a model generalizes and to compare models.

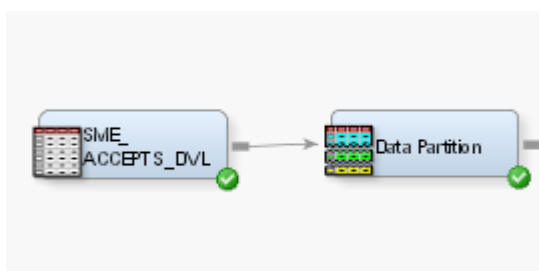


Figure 4.5 Data Partition.

Property	Value
General	
Node ID	Part
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Output Type	Data
Partitioning Method	Stratified
Random Seed	12345
☒ Data Set Allocations	
Training	70.0
Validation	0.0
Test	30.0
Report	
Interval Targets	Yes
Class Targets	Yes
Status	
Create Time	31.03.2012 16:17
Run Id	76bf2000-7b3e-11e1-800
Last Error	
Last Status	Complete
Last Run Time	31.03.2012 17:33
Run Duration	0 Hr. 0 Min. 1,77 Sec.
Grid Host	

Figure 4.6 Data Partition Node Properties Window.

Table 4.5 Data Partition Node Result.

Data Set	Good		Bad	
	Count	Percent	Count	Percent
Development Sample	13209	78%	3683	22%
Training Data	9245	78%	2577	22%
Test Data	3964	78%	1106	22%

4.5.2 Initial Characteristic Analysis

After the determination of the data set to be used in model construction, data analysis phase begins. However, characteristics in the data set are transformed in order

to be able to put the data in more usable format and to make the information value of the characteristics more comprehensible before starting to statistical data analysis phase (Anderson, 2007).

In order to display high performance, there must be linear association between the dependent variable and predictor variable in the linear models such as logistic regression which is the most commonly preferred model in construction of credit scoring model.

Hence, in the case of the predictor variable has non linear relationship with the dependent variable the predictor variable is transformed.

There are mainly two methodologies used for transformation named Univariate Methodology and Bivariate Methodology.

4.5.2.1 Univariate Methodology

In this method, the distribution of independent variable is checked whether it fits the normal distribution or not without considering its relationship with the dependent variable. In the cases where the independent variable is not distributed normally, univariate transformation algorithms such as logarithmic, exponential and z-standardization are used to make the predictor variable distributed normally (Falkenstein, 2002).

For the reasons outlined below, the use of this methodology in credit scoring models is rare:

1. It is impossible to represent some non-linear relationship by using univariate transformation techniques.
2. Some important information value of the variable is lost by transformation.
3. It is quite operationally difficult to integrate univariate transformation algorithms into application system of the banks.

4.5.2.2 Bivariate Methodology

Unlike the univariate method, the relationship between the dependent variable and independent variable is taken into consideration while transforming the independent variable by bivariate transformation method. Dummy variable and the weight of evidence are the most often used algorithms of bivariate transformation method.

4.5.2.2.1 Dummy Variable Transformation

In dummy transformation method, the original value of the predictor variable is replaced with the dummy variable which consists of a set of 2 values that are generally represented with 0 and 1 and is binary. Dummy variable can take only one of these values.

4.5.2.2.2 Weight of Evidence Transformation

Another method used to transform the independent variable is using the weight of evidence (WOE), whose calculation algorithm is formulated below, rather than using the actual value of the independent.

$$WoE = Ln \left(\frac{g_i / G}{b_i / B} \right) \quad (4.1)$$

where :

g_i is the number of goods within the interval.

b_i is the number of bads within the interval.

G is the total number of goods in the whole population.

B is the number of bads in the whole population.

Any interval that has a higher proportion of goods than average will have a positive weight of evidence which means less risky class in terms of credit risk and any interval that has a higher than average proportion of bads will have a negative WoE which means risky class in terms of credit risk.

Initial characteristic analysis involves two main tasks. The first step is to assess the strength of each characteristic individually as a predictor of performance. This is also known as univariate screening, and is done to screen out weak or illogical characteristics.

The strongest characteristics are then grouped. This applies to attributes in both continuous and discrete characteristics.

A large number of attributes are derived as grouping the variables defined in the data dictionary. For instance, in a data set consisting 40 variables, if each variable is grouped into 5 classes, then it results in approximately a data set consisting 200 variables.

To reduce the complexity, characteristic selection procedure is applied. In general, the aim of characteristic selection can be categorized into 3 headlines (Guyon and Elisseeff, 2003):

1. Increasing the prediction power of the characteristics
2. Providing a faster and cost-effective independent variables
3. Reducing the issues resulted from highly correlative variables

4.5.3 Statistical Measures Used to Characteristic Selection

Characteristic selection is often an essential data processing step prior to applying a learning algorithm. The removal of irrelevant and redundant information often improves the performance of predictive models (Hall and Smith, 1998).

Having a great number of non predictive variables or variables with weak prediction power within the modeling stage can result in decrease the predictive power or the final model. In this section, the statistical methods, that can be used in order to detect and eliminate these kinds of variables, are mentioned.

4.5.3.1 Information Value

While constructing the classification models, Information Value (IV) is one of the most commonly used statistical measure to gauge the association power between the dependent and predictor variable (Anderson, 2007).

IV for a classed variable is calculated with the equation shown below:

$$IV = \sum_{i=1}^I \left(\frac{g_i}{G} - \frac{b_i}{B} \right) \ln \left(\frac{g_i / G}{b_i / B} \right) \quad (4.2)$$

where:

I is the number of intervals that the variable has been classed into.

g_i is the number of goods in the i^{th} interval.

b_i is the number of bads in the i^{th} interval.

G is the number of goods in the whole sample.

B is the number of bads in the whole sample.

Table 4.6 Information value (F-statistic) Benchmarks.

Value Range	Strength	Description
$F \leq 0.01$	Uncorrelated	Drop, unless sound reason to consider further.
$0.01 < F \leq 0.05$	Very weak	May add value, but can be dropped with minimal loss.
$0.05 < F \leq 0.10$	Weak	Possible spurious correlation, unlikely to feature.
$0.10 < F \leq 0.30$	Medium	Known correlation, which could appear.
$0.30 < F \leq 0.50$	Strong	High information content, and sought after.
$0.50 < F \leq 1.00$	Dominant!	Very powerful, but investigate why.
$1.00 < F$	Warning!	Possible error, perhaps outcome posing as observation.

4.5.3.2 Chi-Square

Chi-square statistic (χ^2) is a metric that measures the magnitude of the association between the dependent variable and predictor variable by using the distribution of the classes of the predictor variable. Chi-square statistic can take values between 0 and infinity and is calculated as

$$\text{Chi - squared statistics}(\chi^2) = \sum_{i=1}^I \left(\frac{\hat{g}_i - g_i}{\hat{g}_i} \right)^2 + \left(\frac{\hat{b}_i - b_i}{\hat{b}_i} \right)^2 \quad (4.3)$$

where:

I is the number of intervals that the variable has been classed into.

g_i is the number of goods in the i^{th} interval.

b_i is the number of bads in the i^{th} interval.

$\hat{g}_i$ is the expected number of goods in the interval if the distribution was completely uniform.

$\hat{b}_i$ is the expected number of bads in the interval if the distribution was completely uniform.

G is the number of goods in the whole sample.

B is the number of bads in the whole sample.

where:

$$\hat{g}_i = \frac{(g_i + b_i)G}{G + B} \quad (4.4)$$

$$\hat{b}_i = \frac{(g_i + b_i)G}{G + B} \quad (4.5)$$

Table 4.7 Chi-Square (X^2) Benchmarks.

Value Range	Strength
$30 < X^2 \leq 150$	Moderate
$150 < X^2$	Strong

4.5.3.3 Gini Coefficient

Another metric used to measure the prediction power of a grouped variable is Gini Coefficient which is also known as Somer's D Concordance Statistic. In this method, firstly all of the groups for a given variable is ranked in accordance of bad rate from high to low then Gini is calculated by the formula shown below:

$$Gini = 1 - \sum_{i=2}^I [G(i) + G(i-1)] [B(i) - B(i-1)] \quad (4.6)$$

where:

I is the number of intervals that the variable has been classed into.

G(i) is the cumulative proportion of goods in the i^{th} interval.

$B(i)$ is the cumulative proportion of bads in the i^{th} interval.

Table 4.8 Gini Benchmarks.

Value Range	Strength
Gini > 0.05	Moderate
0.25 < Gini	Strong

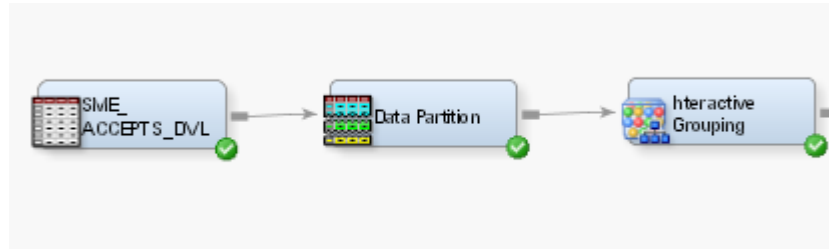


Figure 4.7 Interactive Grouping Node

In this application, Gini Statistic is used as variable selection method. Variables with Gini statistic greater than the specified cutoff, here in this application Gini Cutoff is set to 5.0, will be selected as inputs to successor nodes. Otherwise, the variables will have a role of "Rejected".

Property	Value
Interactive Grouping	
Interval Variable Binning Options	
Binning Method	Quantile
Number of Bins	20
Special Code Options	
Use Special Codes	No
Special Codes Data Set	
Grouping Options	
Use Frozen Groupings	Yes
Criterion	Entropy
Missing Values	Separate Branch if Any
Interval Grouping Method	Optimal
Ordinal Grouping Method	Optimal
Maximum Branch	5
Minimum Categorical Size	5
Significant Digits	2
Apply Restrictions	Yes
Type	Percent
Percent	5.0
Count	1
Adjust WOE	Yes
Adjustment Factor	0.5
Import/Export Options	
Import Grouping Data	No
Import Dataset	
Score	
Group Level	Ordinal
Variable Selection Method	Gini Statistic
Gini Cutoff	5.0
Information Value Cutoff	0.1
Report	
Create Grouping Data	Yes
Create Method	Overwrite
Number of Variables	10
Status	

Figure 4.8 Interactive Grouping Node Properties Window.

Table 4.9 Interactive Grouping Node Output.

Variable	Group	Group Values	ROLE	Gini Ordering	GINI	Weight of Evidence	Event Rate	Non Event Rate
cc_stat	1	K, N, T	Input	1	37.8577183	-4.478847444	96.0886571	3.911342894
cc_stat	2	F, L, S	Input	1	37.8577183	0.227523398	18.16839	81.83161004
cc_stat	3	A, G, X, Missing	Input	1	37.8577183	0.188114663	18.7616643	81.23833571
cc_stat	4	E, M, O, W	Input	1	37.8577183	3.782761861	0.63041765	99.36958235
num_empl	1	6, 7	Input	2	15.2074106	-0.626510107	34.2776204	65.7223796
num_empl	2	10, 8, 9	Input	2	15.2074106	-0.274726416	26.8404908	73.1595092
num_empl	3	2, 4	Input	2	15.2074106	-0.02654137	22.2541691	77.74583094
num_empl	4	1, 11, 5, Missing	Input	2	15.2074106	0.15102122	19.3335916	80.66640837
num_empl	5	3	Input	2	15.2074106	0.789680967	11.2332112	88.76678877
firm_age	1	1, 10, 4, 9	Input	3	15.0546804	0.904563481	10.1376721	89.86232791
firm_age	2	5	Input	3	15.0546804	0.525352372	14.1509434	85.8490566
firm_age	3	2, Missing	Input	3	15.0546804	-0.094415965	23.4505863	76.54941374
firm_age	4	3, 8	Input	3	15.0546804	-0.175924934	24.9452954	75.0547046
firm_age	5	11, 6, 7	Input	3	15.0546804	-0.511989091	31.7460317	68.25396825
avg_term	1	41, 45, 46	Input	4	14.3028156	-0.248992918	26.3381995	73.66180049
avg_term	2	40, 43, 53	Input	4	14.3028156	-0.170009104	24.8346998	75.16530019
avg_term	3	44, 51, 55, 56, 57, 58, 59	Input	4	14.3028156	0.072194991	20.5927284	79.40727162
avg_term	4	42, 49, 52, 54	Input	4	14.3028156	0.987277936	9.40860215	90.59139785
avg_term	5	48, 50, Missing	Input	4	14.3028156	-0.119273684	23.8997555	76.1002445
Age	1	Age< 2406	Input	5	13.4520437	-0.542912384	32.4198251	67.58017493
Age	2	2406<= Age< 3009	Input	5	13.4520437	-0.04251818	22.5318197	77.46818025
Age	3	3009<= Age< 3603	Input	5	13.4520437	0.103162922	20.0909467	79.90905333
Age	4	3603<= Age< 4907, Missing	Input	5	13.4520437	0.168858991	19.0569196	80.94308036

Table 4.9 (continued)

Variable	Group	Group Values	ROLE	Gini Ordering	GINI	Weight of Evidence	Event Rate	Non Event Rate
Age	5	4907<= Age	Input	5	13.4520437	0.347094617	16.4578112	83.54218881
cus_tenure	1	0	Input	6	12.9623812	0.034376784	21.2180159	78.78198414
cus_tenure	2	1	Input	6	12.9623812	0.072991296	20.5797101	79.42028986
cus_tenure	3	2, 3, 4	Input	6	12.9623812	0.448309167	15.1129944	84.88700565
cus_tenure	4	10, 11, 12, 5, 6, 7, 8, 9	Input	6	12.9623812	1.088529671	8.58034321	91.41965679
cus_tenure	5	13, Missing	Input	6	12.9623812	-0.312097602	27.5806452	72.41935484
dom_sales_con	1	dom_sales_con< 306	Input	7	12.2544756	0.045192163	21.0377891	78.96221094
dom_sales_con	2	306<= dom_sales_con < 314, Missing	Input	7	12.2544756	-0.138182025	24.2453514	75.75464864
dom_sales_con	3	314<= dom_sales_con < 342	Input	7	12.2544756	0.454090685	15.0389729	84.96102705
dom_sales_con	4	342<= dom_sales_con < 353	Input	7	12.2544756	0.055209632	20.8718626	79.12813738
dom_sales_con	5	353<= dom_sales_con	Input	7	12.2544756	-0.357069034	28.4878864	71.51211362
sector	1	sector< 212	Input	8	10.8761052	0.159162512	19.2069393	80.79306072
sector	2	212<= sector< 217	Input	8	10.8761052	0.555878963	13.7841352	86.21586476
sector	3	217<= sector< 229	Input	8	10.8761052	-0.28682115	27.0786517	72.92134831
sector	4	229<= sector	Input	8	10.8761052	-0.060015912	22.8387097	77.16129032
sector	5	Missing	Input	8	10.8761052	0.057037933	20.8416834	79.15831663
existing_cus	1	H	Input	9	8.93798429	-0.313880928	27.6162791	72.38372093
existing_cus	2	E, Missing	Input	9	8.93798429	0.125538075	19.7341279	80.26587211
msh_oth_inc	1	msh_oth_inc < 1, Missing	Input	10	8.81767468	0.144399713	19.4370685	80.56293149
msh_oth_inc	2	1<= msh_oth_inc < 1200	Input	10	8.81767468	-0.121247153	23.935667	76.06433302
msh_oth_inc	3	1200<= msh_oth_inc < 2500	Input	10	8.81767468	0.142558458	19.4659171	80.53408292
msh_oth_inc	4	2500<= msh_oth_inc < 6000	Input	10	8.81767468	-0.326074881	27.8606965	72.13930348
msh_oth_inc	5	6000<= msh_oth_inc	Input	10	8.81767468	-0.020620639	22.1518987	77.84810127
ho_kk_limit	1	ho_kk_limit< 1000, Missing	Input	11	8.4874791	-0.100890899	23.5670194	76.4329806

Table 4.9 (continued)

Variable	Group	Group Values	ROLE	Gini Ordering	GINI	Weight of Evidence	Event Rate	Non Event Rate
msh_cc_limit	2	1000<= msh_cc_limit< 1600	Input	11	8.4874791	0.175033095	18.9618644	81.03813559
msh_cc_limit	3	1600msh_cc_limit< 5000	Input	11	8.4874791	0.712578021	12.0253165	87.97468354
msh_cc_limit	4	5000<= msh_cc_limit	Input	11	8.4874791	0.229241697	18.1428571	81.85714286
cus_type	1	DM, FU, KS, MB, MT, MU, SM, Missing	Input	12	8.35811154	-0.104708629	23.6358575	76.36414254
cus_type	2	MS, MY	Input	12	8.35811154	0.001903348	21.7659138	78.23408624
cus_type	3	MO	Input	12	8.35811154	0.079423543	20.4747774	79.52522255
cus_type	4	JP, MA	Input	12	8.35811154	0.220954726	18.2662539	81.73374613
cus_type	5	JM, JV, P	Input	12	8.35811154	0.861976519	10.5322763	89.46772367
spouse_work	1	H	Input	13	6.83799127	0.068089047	20.6599504	79.34004964
spouse_work	2	E	Input	13	6.83799127	0.390769053	15.8660844	84.13391557
spouse_work	3	Missing	Input	13	6.83799127	-0.155689668	24.5683621	75.43163789
firm_web	1	H, Missing	Input	14	5.12139148	-0.065197183	22.9301458	77.06985415
firm_web	2	E	Input	14	5.12139148	0.240244653	17.9800222	82.0199778
gender	1	K	Rejected	15	4.9301293	-0.233976715	26.0479042	73.95209581
gender	2	E, Missing	Rejected	15	4.9301293	0.06255935	20.7507381	79.24926191
supplier_risk	1	H	Rejected	16	4.40174586	-0.08153035	23.2200647	76.77993528
supplier_risk	2	E	Rejected	16	4.40174586	0.094493146	20.2304965	79.76950355
supplier_risk	3	Missing	Rejected	16	4.40174586	-1.277456933	50	50
firm_city_code	1	firm_city_code< 7	Rejected	17	4.39692726	0.077984386	20.4982206	79.50177936
firm_city_code	2	7<= firm_city_code< 16	Rejected	17	4.39692726	-0.243006589	26.2222222	73.77777778
firm_city_code	3	16<= firm_city_code< 42	Rejected	17	4.39692726	-0.02064947	22.1523959	77.84760408
firm_city_code	4	42<= firm_city_code	Rejected	17	4.39692726	0.112283161	19.9449253	80.05507474
firm_city_code	5	Missing	Rejected	17	4.39692726	-0.088988315	23.3532934	76.64670659
mgmt_risk	1	H, Missing	Rejected	18	4.35986437	-0.08153035	23.2200647	76.77993528

Table 4.9 (continued)

Variable	Group	Group Values	ROLE	Gini Ordering	_GINI_	Weight of Evidence	Event Rate	Non Event Rate
mgmt_risk	2	E	Rejected	18	4.35986437	0.093839352	20.2410493	79.75895073
payment_type	1	K	Rejected	19	2.3669718	-0.212963092	25.6451613	74.35483871
payment_type	2	P, Missing	Rejected	19	2.3669718	0.026643596	21.3475713	78.65242865
id_type	1	N	Rejected	20	1.23029932	-0.012543147	22.0129171	77.98708288
id_type	2	E, P	Rejected	20	1.23029932	0.256290159	17.7446103	82.25538972
id_type	3	Missing	Rejected	20	1.23029932	-0.04271247	22.5352113	77.46478873
drwn_cheque	1	H, Missing	Rejected	21	0.66222122	-0.008437982	21.9425238	78.05747623
drwn_cheque	2	E	Rejected	21	0.66222122	0.030776881	21.2782541	78.72174591
maritual_stat	1	B	Rejected	22	0	-0.302694433	27.3932253	72.60677467
maritual_stat	2	E, Missing	Rejected	22	0	0	0	0
msh_oto	NaN		Rejected	23	NaN	NaN	NaN	NaN
real_est_rec	NaN		Rejected	24	NaN	NaN	NaN	NaN
col_prvd	NaN		Rejected	25	NaN	NaN	NaN	NaN
msh_member_org	NaN		Rejected	26	NaN	NaN	NaN	NaN
nationality	NaN		Rejected	27	NaN	NaN	NaN	NaN

Weight of Evidence (WOE) measures the relative risk of an attribute or group level. The value depends on the value of the binary target variable, which is either event or nonevent. In the following equations, n_i^{event} and n_i^{nonevent} denote the numbers of events and nonevents in the attribute. N^{event} and N^{nonevent} represent the total numbers of events and nonevents in the data. An attribute's WOE is defined as follows:

$$\text{Weight of Evidence}_{\text{attribute}} = \ln \frac{P(\text{Good})_{\text{attribute}}}{P(\text{Bad})_{\text{attribute}}} \quad (4.7)$$

where:

$$P(\text{Good})_{\text{attribute}} = \ln \frac{\text{number of good}_{\text{attribute}}}{\text{number of good}_{\text{total}}} \quad (4.8)$$

$$P(\text{Bad})_{\text{attribute}} = \ln \frac{\text{number of bad}_{\text{attribute}}}{\text{number of bad}_{\text{total}}} \quad (4.9)$$

The Table 5.7 shows a typical chart used in analysis of grouped characteristics. The example shows the characteristic "cc_stat" after it has been grouped.

Table 4.10 Analysis of Grouped Variables.

Group Values	Count	EVENT COUNT	NONEVENTCOUNT	Distr Good	Distr Bad	Weight of Evidence
K, N, T	767	737	30	0.003	0.286	-4.478847444
F, L, S	677	123	554	0.06	0.048	0.227523398
A, G, X, Missing	9109	1709	7400	0.8	0.663	0.188114663
E, M, O, W	1269	8	1261	0.136	0.003	3.782761861
TOTAL	11822	2577	9245	1	1	

For example, calculation of the WOE of attribute "K,N,T" is:

$$\ln(\text{Distr Good} / \text{Distr Bad}) = \ln(0.003/0.286) = -4.478847444$$

The WOE is a standardized way of comparing the interval Good:Bad odds with the average Good:Bad odds of the sample. Any interval that has a higher proportion of

goods than average will have a positive weight of evidence, and any interval that has a higher than average proportion of bads will have a negative WOE. A WOE of 1 means the good:bad odds are e^1 (2.718) times average, a value of $2 e^2$ (7.389) times average and so on. Likewise, a value of -1 means the odds are 2.718 times smaller than average.

4.6 MODEL BUILDING

Logistic regression is a common technique used to develop scorecards in most financial industry applications, where the predicted variable is categorical.

In this application, first a regression model is built to calculate score then this score is transformed into probability.

$$f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (4.10)$$

where :

$f(x)$ = score

β_0 = intercept of the regression line

β_n = parameters

X_n = input variables

$$P(Y = 1) = \frac{e^{f(x)}}{1 + e^{f(x)}} \quad (4.11)$$

where :

e : takes a value of 2.7183

$P(Y=1)$: Probability of an event occurs. In this application, this implies the probability of an application is default. So we can change the formula as shown below:

$$P(GB = 1) = \frac{e^{f(x)}}{1 + e^{f(x)}} \quad (4.12)$$



Figure 4.9 Regression Node.

The statistical modeling technique employed for this development is Logistic Regression. In this application Stepwise method is used as selection model. This method begins, by default, with no candidate effects in the model and then systematically adds effects that are significantly associated with the target. However, after an effect is added to the model, Stepwise may remove any effect already in the model that is not significantly associated with the target. This stepwise process continues until one of the following occurs:

- No other effect in the model meets the Stay Significance Level.
- The Max Steps criterion is met. If you choose the Stepwise selection method, then you can specify a Max Steps to put a limit on the number of steps before the effect selection process stops. The default value is set to the number of effects in the model. If you add interactions via the Interaction Builder, the Max Steps is automatically updated to include these terms.
- An effect added in one step is the only effect deleted in the next step.

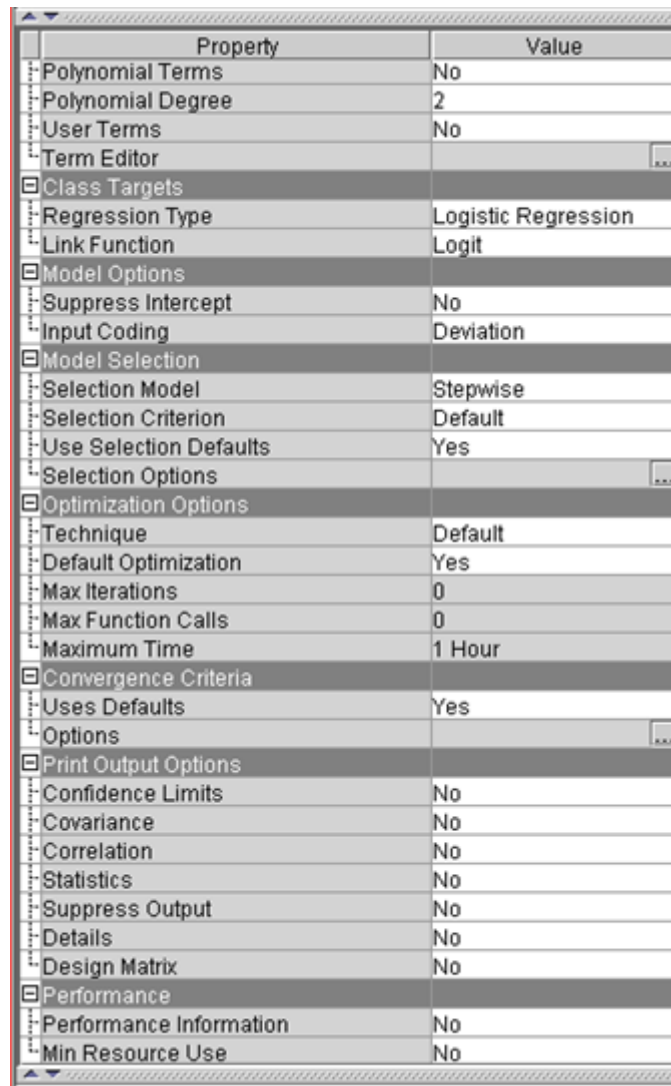


Figure 4.10 Regression Node Property Window.

A useful property of WOE is that it always displays a linear relationship with the natural log of the good:bad odds ($\ln(\text{odds})$). This is very useful for pre-processing data, particularly when linear models are being constructed. Because of that reason, in this application, WOE values of the characteristics are used instead of using characteristics itself. The predictors that are used as input to regression model are shown below:

Table 4.11 Regression Model Inputs.

Name	Use	Role	Level
GB	Yes	TARGET	BINARY
WOE_dom_sales_con	Yes	INPUT	INTERVAL
WOE_sector	Yes	INPUT	INTERVAL
WOE_num_empl	Yes	INPUT	INTERVAL
WOE_avg_term	Yes	INPUT	INTERVAL
WOE_cus_type	Yes	INPUT	INTERVAL

Table 4.11 (continued)

Name	Use	Role	Level
WOE_cc_stat	Yes	INPUT	INTERVAL
WOE_msh_cc_limit	Yes	INPUT	INTERVAL
WOE_msh_oth_inc	Yes	INPUT	INTERVAL
WOE_firm_web	Yes	INPUT	INTERVAL
WOE_firm_age	Yes	INPUT	INTERVAL
WOE_spouse_work	Yes	INPUT	INTERVAL
WOE_cus_tenure	Yes	INPUT	INTERVAL
WOE_existing_cus	Yes	INPUT	INTERVAL
WOE_age	Yes	INPUT	INTERVAL

Table 4.12 Regression Model Output.

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	Standardized Estimate	Exp(Est)
Intercept	1	-1.2634	0.0271	2167.2	<.0001	0.283	
WOE_age	1	-0.9365	0.1018	84.64	<.0001	-0.1325	0.392
WOE_existing_cus	1	0.5193	0.2023	6.59	0.0103	0.0553	1.681
WOE_cus_tenure	1	-0.6709	0.145	21.41	<.0001	-0.1182	0.511
WOE_firm_age	1	-0.4545	0.0977	21.65	<.0001	-0.0974	0.635
WOE_firm_web	1	-1.1076	0.2233	24.61	<.0001	-0.0783	0.33
WOE_msh_oth_inc	1	-0.4107	0.1634	6.32	0.0119	-0.038	0.663
WOE_msh_cc_limit	1	-0.6693	0.1332	25.24	<.0001	-0.0908	0.512
WOE_cc-stat	1	-1.0024	0.0376	709.8	<.0001	-0.9237	0.367
WOE_avg_term	1	-0.4492	0.1032	18.93	<.0001	-0.0935	0.638
WOE_num_empl	1	-0.7689	0.0979	61.66	<.0001	-0.1315	0.464
WOE_dom_sales_con	1	-0.3236	0.1203	7.23	0.0072	-0.043	0.724

In this application, logistic regression model is used to estimate credit worthy of applicant (one who applies for the loan) in terms of probability of default. In logistic regression model, here in this application takes name of "PD Application Scorecard Model", as the value of regression model function increases, the probability of default increases.

In order to make the relationship between dependent variable and predictor variables linear, sensible and significant, weight of evidence (WOE) value of the predictor variables are used in modeling instead of using raw value of the predictor

variables. Any group of the predictor variable that has a higher proportion of goods than average will have a positive WOE, and any group of the predictor variable that has a higher proportion of bads than average will have a negative WOE. In PD Application Scorecard Models, groups with positive WOE means less risky group in terms of credit risk, and groups with negative WOE means risky group in terms of credit risk.

For a value of predictor variable which falls into a risky group, then the WOE will have negative sign. In order to ensure the necessity of "as the value of regression model function increases, the probability of default increases" model coefficient of the predictor variable must have negative sign.

Hence the sign of the coefficient of "WOE_existing_cus" characteristic is positive, there is correlation among the variables which are selected as input to regression model. Therefore, correlation analysis should be performed and the variables with a high correlation should be removed from the list of variables that will be put into regression model as input.

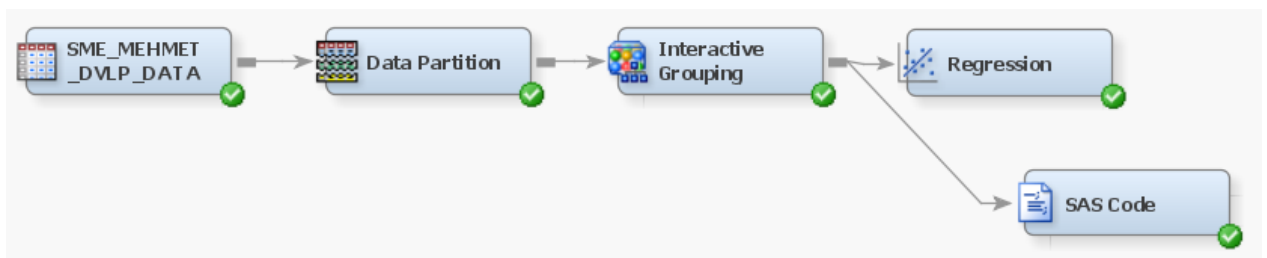


Figure 4.11 SAS Code Node.

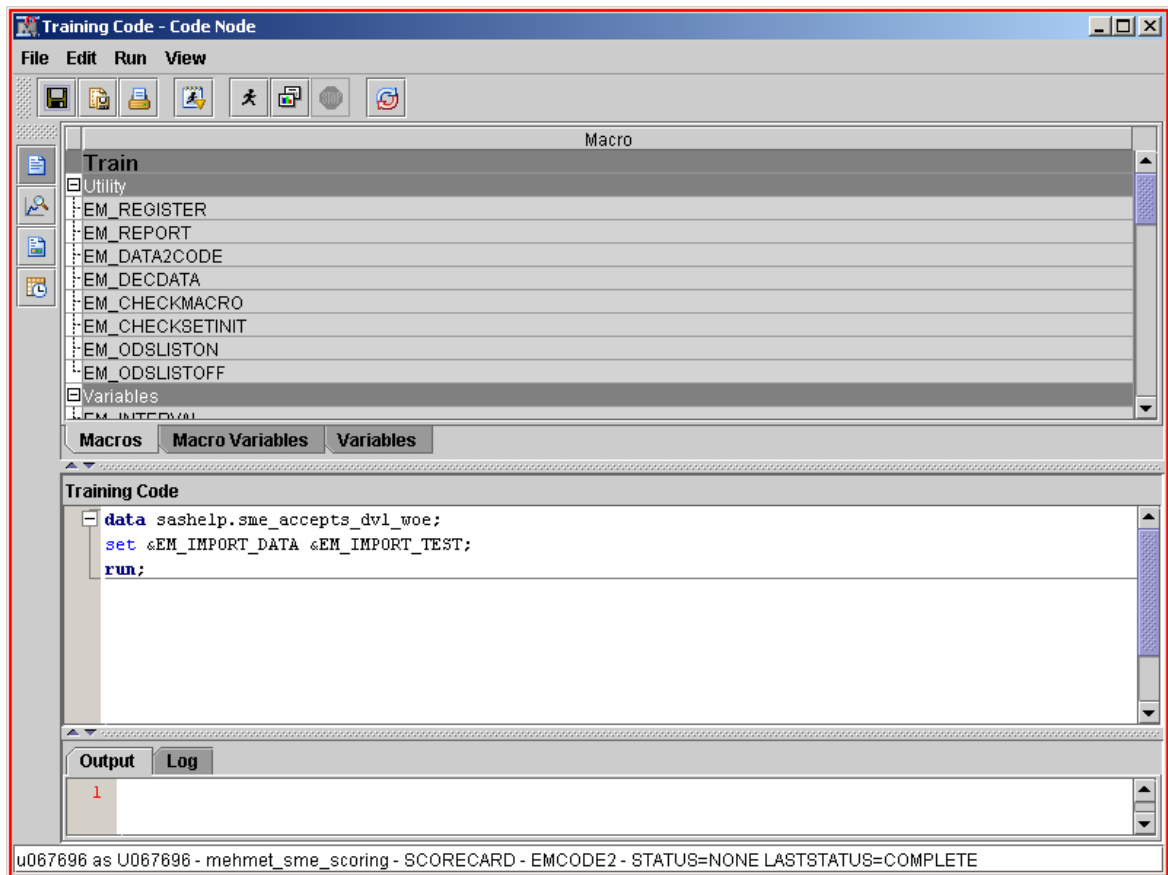


Figure 4.12 SAS Code Editor Window.

In order to perform correlation analysis, both train data and test data, which were produced at stage of data partition, are transferred into SAS Enterprise Guide by using SAS Code which is shown on the window above.

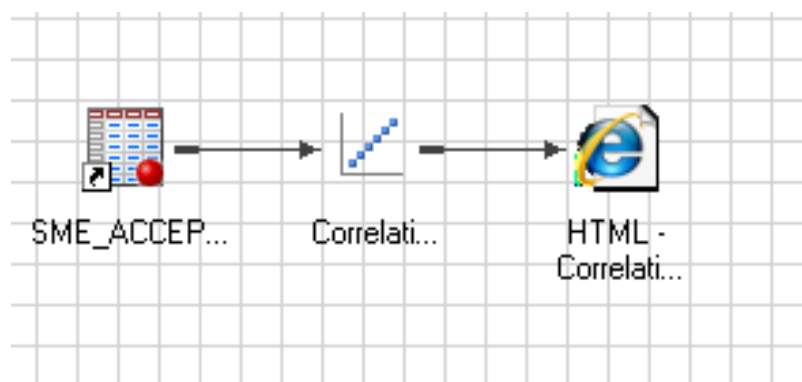


Figure 4.13 Correlation Analysis Diagram on SAS Enterprise Guide.

4.6.1 Pearson Correlation Analysis

Table 4.13 Correlation Analysis Output.

Characteristic	First Correlative Characteristic	Second Correlative Characteristic
WOE_dom_sales_con	WOE_dom_sales_con	WOE_avg_term
Weight of Evidence: dom_sales_con	1	0.31621
WOE_sector	WOE_sector	WOE_avg_term
Weight of Evidence: sector	1	0.43546
WOE_num_empl	WOE_num_empl	WOE_cus_type
Weight of Evidence: num_empl	1	0.63873
WOE_avg_term	WOE_avg_term	WOE_firm_age
Weight of Evidence: avg_term	1	0.59435
WOE_cus_type	WOE_cus_type	WOE_num_empl
Weight of Evidence: cus_type	1	0.63873
WOE_cc_stat	WOE_cc_stat	WOE_msh_oth_inc
Weight of Evidence: cc_stat	1	0.10562
WOE_msh_cc_limit	WOE_msh_cc_limit	WOE_firm_age
Weight of Evidence:msh_cc_limit	1	0.19916
WOE_msh_oth_inc	WOE_msh_oth_inc	WOE_existing_cus
Weight of Evidence: msh_oth_inc	1	0.15191
WOE_firm_web	WOE_firm_web	WOE_avg_term
Weight of Evidence: firm_web	1	0.15439
WOE_firm_age	WOE_firm_age	WOE_avg_term
Weight of Evidence: firm_age	1	0.59435
WOE_spouse_work	WOE_spouse_work	WOE_cus_tenure
Weight of Evidence:spouse_work	1	0.52089
WOE_cus_tenure	WOE_cus_tenure	WOE_existing_cus
Weight of Evidence: cus_tenure	1	0.6367
WOE_existing_cus	WOE_existing_cus	WOE_cus_tenure
Weight of Evidence: existing_cus	1	0.6367
WOE_Age	WOE_Age	WOE_firm_age
Weight of Evidence: Age	1	0.20867

From the table, correlation between WOE_existing_cus and WOE_cus_tenure is the highest so one of them should be removed. Since WOE of cus_tenure is greater than WOE of existing_cus, WOE_existing_cus is omitted.

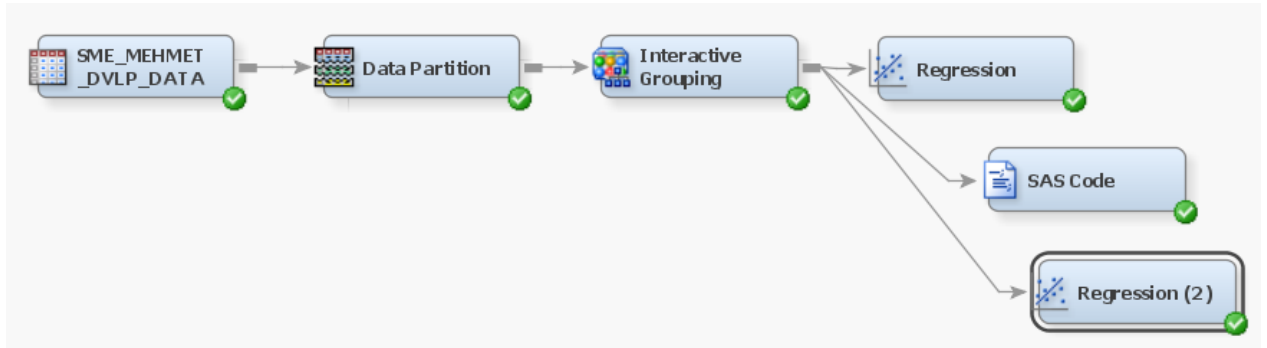


Figure 4.14 Regression Node(2).

After performing correlation analysis Regression(2) node connects to Interactive Grouping node. The role of WOE_existing_cus is set to REJECTED as shown Table 4.3.4 in order to prevent entering Regression since it was measured as the highest correlative characteristics at the correlation analysis stage.

Table 4.14 Characteristics and Their Roles.

Name	Use	Report	Role	Level
GB	Yes	N	TARGET	BINARY
WOE_Age	Yes	N	INPUT	INTERVAL
WOE_existing_cus	No	N	REJECTED	INTERVAL
WOE_cus_tenure	Yes	N	INPUT	INTERVAL
WOE_spouse_work	Yes	N	INPUT	INTERVAL
WOE_firm_age	Yes	N	INPUT	INTERVAL
WOE_firm_web	Yes	N	INPUT	INTERVAL
WOE_msh_oth_inc	Yes	N	INPUT	INTERVAL
WOE_cc_limit	Yes	N	INPUT	INTERVAL
WOE_cc_stat	Yes	N	INPUT	INTERVAL
WOE_cus_type	Yes	N	INPUT	INTERVAL
WOE_avg_term	Yes	N	INPUT	INTERVAL
WOE_num_empl	Yes	N	INPUT	INTERVAL
WOE_sector	Yes	N	INPUT	INTERVAL
WOE_dom_sales_con	Yes	N	INPUT	INTERVAL

Table 4.14 Summary of Stepwise Selection.

Step	Effect Entered	DF			Pr>ChiSq
1	WOE_cc_stat	1	1	2335.7427	<.0001
2	WOE_firm_age	1	2	181.2994	<.0001
3	WOE_num_empl	1	3	136.8499	<.0001
4	WOE_Age	1	4	89.1594	<.0001
5	WOE_firm_web	1	5	40.1004	<.0001
6	WOE_avg_term	1	6	30.0326	<.0001
7	WOE_cc_limit	1	7	21.247	<.0001
8	WOE_cus_tenure	1	8	19.5455	<.0001
9	WOE_som_sales_con	1	9	7.5426	0.006
10	WOE_msh_oth_inc	1	10	4.8983	0.0269

The selected model, based on the CHOOSE=NONE criterion, is the model trained in Step 10. It consists of the following effects:

Intercept , WOE_Age, WOE_cus_tenure, WOE_firm_age , WOE_firm_web , WOE_msh_oth_inc , WOE_cc_limit, WOE_cc_stat , WOE_avg_term, WOE_num_empl, WOE_dom_sales_con.

Table 4.15 Regression Model Elements.

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	Standardized Estimate	Exp(Est)
Intercept	1	-1.2627	0.0271	2170.46	<.0001	0.283	
WOE_Age	1	-0.9823	0.1003	96	<.0001	-0.139	0.374
WOE_cus_tenure	1	-0.4236	0.1015	17.42	<.0001	-0.0746	0.655
WOE_firm_age	1	-0.4567	0.0975	21.96	<.0001	-0.0979	0.633
WOE_firm_web	1	-1.2033	0.2202	29.87	<.0001	-0.0851	0.3
WOE_msh_oth_inc	1	-0.3588	0.1622	4.9	0.0269	-0.0332	0.698
WOE_cc_limit	1	-0.6499	0.1327	23.98	<.0001	-0.0882	0.522
WOE_cc_stat	1	-1.0002	0.0376	707.69	<.0001	-0.9216	0.368
WOE_avg_term	1	-0.463	0.1032	20.14	<.0001	-0.0963	0.629
WOE_num_empl	1	-0.7094	0.0945	56.37	<.0001	-0.1213	0.492
WOE_dom_sales_con	1	-0.3107	0.1199	6.71	0.0096	-0.0413	0.733

4.6.2 Misclassification

Credit Scorecards are designed to predict to the probability of an applicant being good or bad. They are used for differentiating between good and bad applications. Misclassification statistics are a good way to determine whether a scorecard is providing the right differentiation.

		Predicted	
		Good	Bad
Actual	Good	True Positive	False Negative
	Bad	False Positive	True Negative

Figure 4.15 Confusion Matrix.

True Positive : Acceptance of goods

False Positive : Acceptance of bads

False Negative: Decline goods

True Negative : Decline bads

Measures to gauge misclassification:

Accuracy : $(\text{True Positive} + \text{True Negative}) / \text{total cases}$

Error Rate : $(\text{False Positive} + \text{False Negative}) / \text{total cases}$

Sensitivity : $\text{True Positive} / \text{total actual positive}$

Specificity : $\text{True Negative} / \text{total actual negative}$

Table 4.16 Confusion Matrix Regression Output.

False Negative	True Negative	False Positive	True Positive
1834	9208	37	743

4.7 MASTER SCORECARD

Table 4.17 Logistic Regression Output.

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	Standardized Estimate	Exp(Est)
Intercept	1	-1.2627	0.0271	2170.46	<.0001	0.283	
WOE_Age	1	-0.9823	0.1003	96	<.0001	-0.139	0.374
WOE_cus_tenure	1	-0.4236	0.1015	17.42	<.0001	-0.0746	0.655
WOE_firm_age	1	-0.4567	0.0975	21.96	<.0001	-0.0979	0.633
WOE_firm_web	1	-1.2033	0.2202	29.87	<.0001	-0.0851	0.3
WOE_msh_oth_inc	1	-0.3588	0.1622	4.9	0.0269	-0.0332	0.698
WOE_cc_limit	1	-0.6499	0.1327	23.98	<.0001	-0.0882	0.522
WOE_cc_stat	1	-1.0002	0.0376	707.69	<.0001	-0.9216	0.368
WOE_avg_term	1	-0.463	0.1032	20.14	<.0001	-0.0963	0.629
WOE_num_empl	1	-0.7094	0.0945	56.37	<.0001	-0.1213	0.492
WOE_dom_sales_con	1	-0.3107	0.1199	6.71	0.0096	-0.0413	0.733

Since Weights of Evidence (WOE) transformed versions of the predictor variables have been used, the WOE prefix that has been added to the names of the predictor variables. The parameter coefficient for the i^{th} interval within a predictor variable is calculated as:

$$b'_t = \text{WOE}_t * b$$

where:

b'_t is the parameter coefficient of the model associated with the i^{th} coarse classed interval for a given variable. [contribution to the score]

WOE_t is the weight of evidence for the i^{th} coarse classed interval for a given variable.

b is the value of the parameter coefficient assigned to the weight of evidence transformed variable within the model (column named Estimate in Table 5.1).

An example calculation for Age variable is as below:

If the applicant is 25 years old, its WOE becomes -0.04251818 (from column named Weight of Evidence in Table 4.2.1) and then the parameter coefficient for this interval will be $-0.04251818 * -0.9823 = 0.041765608$.

4.7.1 Example Scorecard

Table 4.18 Artificial Variable Value for an Applicant.

Variable	Variable Value for An Applicant
Age	3503
Cus_tenure	5
firm_age	3
firm_web	H
msh_oth_inc	2000
cc_limit	4000
cc_stat	Missing
avg_term	Missing
Num_empl	4
dom_sales_con	200

Table 4.19 Groups and WOE of the Model Parameters.

Variable	Group	Group Values	Weight of Evidence
cc_stat	1	K, N, T	-4.478847444
cc_stat	2	F, L, S	0.227523398
cc_stat	3	A, G, X, Missing	0.188114663
cc_stat	4	E, M, O, W	3.782761861
num_empl	1	6, 7	-0.626510107
num_empl	2	10, 8, 9	-0.274726416
num_empl	3	2, 4	-0.02654137
num_empl	4	1, 11, 5, Missing	0.15102122
num_empl	5	3	0.789680967
firm_age	1	1, 10, 4, 9	0.904563481
firm_age	2	5	0.525352372
firm_age	3	2, Missing	-0.094415965
firm_age	4	3, 8	-0.175924934
firm_age	5	11, 6, 7	-0.511989091
avg_term	1	41, 45, 46	-0.248992918
avg_term	2	40, 43, 53	-0.170009104
avg_term	3	44, 51, 55, 56, 57, 58, 59	0.072194991
avg_term	4	42, 49, 52, 54	0.987277936
avg_term	5	48, 50, Missing	-0.119273684
Age	1	Age < 2406	-0.542912384

Variable	Group	Group Values	Weight of Evidence
Age	2	2406<= Age< 3009	-0.04251818
Age	3	3009<= Age< 3603	0.103162922
Age	4	3603<= Age< 4907, Missing	0.168858991
Age	5	4907<= Age	0.347094617
cus_tenure	1	0	0.034376784
cus_tenure	2	1	0.072991296
cus_tenure	3	2, 3, 4	0.448309167
cus_tenure	4	10, 11, 12, 5, 6, 7, 8, 9	1.088529671
cus_tenure	5	13, Missing	-0.312097602
dom_sales_con	1	yi_sat_kos< 306	0.045192163
dom_sales_con	2	306<= dom_sale_con< 314, Missing	-0.138182025
dom_sales_con	3	314<= dom_sale_con< 342	0.454090685
dom_sales_con	4	342<= dom_sale_con< 353	0.055209632
dom_sales_con	5	353<= dom_sale_con	-0.357069034
msh_oth_inc	1	msh_oth_inc< 1, Missing	0.144399713
msh_oth_inc	2	1<= msh_oth_inc < 1200	-0.121247153
msh_oth_inc	3	1200<= msh_oth_inc< 2500	0.142558458
msh_oth_inc	4	2500<= msh_oth_inc < 6000	-0.326074881
msh_oth_inc	5	6000<= msh_oth_inc	-0.020620639
cc_limit	1	cc_limit < 1000, Missing	-0.100890899
cc_limit	2	1000<= cc_limit < 1600	0.175033095
cc_limit	3	1600<= cc_limit < 5000	0.712578021
cc_limit	4	5000<= cc_limit	0.229241697
firm_web	1	H, Missing	-0.065197183
firm_web	2	E	0.240244653

Table 4.20 Score Calculation.

Variable	Variable Value for An Applicant	WOE	Parameter Coefficient	WOE * Parameter Coefficient
Age	35	0.103162922	-0.9823	-0.101336938
ba_yil	5	1.088529671	-0.4236	-0.461101169
firm_age	3	-0.175924934	-0.4567	0.080344917
firm_web	H	-0.065197183	-1.2033	0.07845177
msh_oth_inc	2000	0.142558458	-0.3588	-0.051149975
cc_limit	4000	0.712578021	-0.6499	-0.463104456
cc_stat	Missing	0.188114663	-1.0002	-0.188152286
avg_term	Missing	-0.119273684	-0.463	0.055223716
num_empl	4	-0.02654137	-0.7094	0.018828448

Variable	Variable Value for An Applicant	WOE	Parameter Coefficient	WOE * Parameter Coefficient
dom_sales_con	200	0.045192163	-0.3107	-0.014041205
TOTAL				-1.0460371772395

$$f(x) = -1.2627 - 1.0460371772395 \Rightarrow f(x) = -2.308737177$$

$$P(GB=1) = \frac{e^{f(x)}}{1 + e^{f(x)}} = \frac{(2.7183)^{(-2.308737177)}}{1 + (2.7183)^{(-2.308737177)}}$$

$$P(GB = 1) = 0.09$$

Credit decision is given by comparing the calculated customer score and Cut-off point which is a score where applicants with scores smaller than this point are accepted and the below this point are rejected. The cut-off score is determined by taking into account several factors such as profit, market strategy, acceptance rate, bad rate expected loss etc. Here in our example, we assume that the cut-off is 0.10. Hence, customer score is smaller than the cut-off value, the credit application is approved.

4.8 MODEL PERFORMANCE

4.8.1 Gini Coefficient

Worldwide accepted standard measure of scorecard performance is the Gini Coefficient. It measures how effectively the scorecard discriminates between ‘Goods’ and ‘Bads’. The Gini Coefficient can take values from 0 to 100. A value of 0 is equivalent to no discrimination while a value of 100 equates to perfect discrimination (which is not practically achievable). It follows that the higher the Gini coefficient the better the scorecard is at discriminating between good and bad customers.

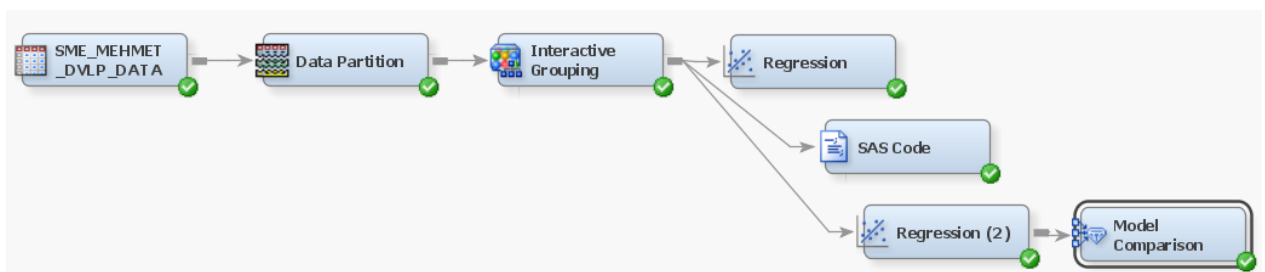


Figure 4.16 Model Comparison Node.

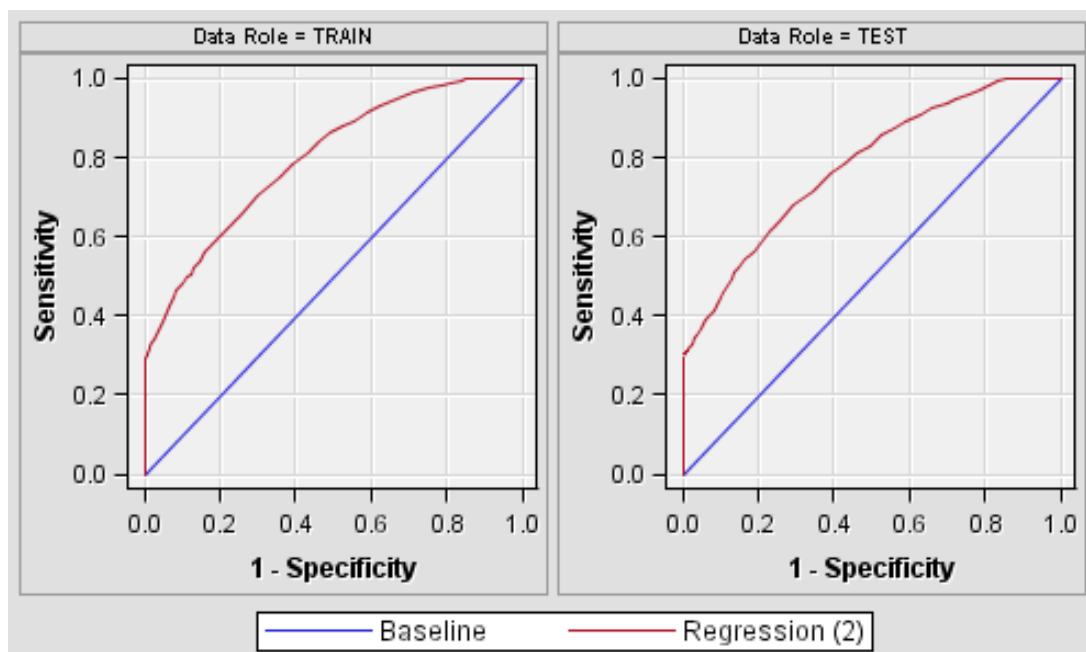


Figure 4.17 ROC for Train and Test Data Samples.

Table 4.21 Model Performance Output.

Target Variable	Data Role	ROC	Gini
GB	Train	0.79	0.59
GB	Test	0.78	0.56

The performance of the model on the test data is very similar to its performance on the training data set. The comparison shows that the model is robust.

Table 4.22 ROC Benchmark Index.

ROC Value	Discrimination Power
$0 < \text{ROC} < 0.5$	No discrimination
$0.5 < \text{ROC} < 0.7$	Poor discrimination
$0.7 < \text{ROC} < 0.8$	Acceptable discrimination
$0.8 < \text{ROC} < 0.9$	Excellent discrimination
$0.9 < \text{ROC} < 1$	Exceptional

This scorecard has a Gini coefficient of 0.59, which proves that the scorecard is strong at discriminating good and bad customers. (rule of thumb for ROC interval)

CHAPTER 5

MANAGERIAL IMPLICATION AND CONCLUSION

A risk exists for decisions taken by banks for almost all business areas. Risk is an integrated part of the banking functions which cannot be prevented and eliminated.

In banking system, the main purpose is to maintain a sustainable and sound credit process system which starts with credit application and ends with closure of the credit. The excellence of this process is closely related to measure and manage the credit risk which is the most critical risk type that banks are exposed.

Various risk factors should be determined carefully in order to be able to control effectively risks that are tied to credit transactions. In today's highly complex banking sector, the measurement of non-payment back, or determination of applicant's credibility should be done carefully, fast, accurate and in realistic way. Thus, efficiency of the crediting function and the rate of responsiveness to the customer needs will increase, and it will be enabled to evaluate more quickly the applications of more customers. The misjudgment of the credit applications by the banks will result in inefficient use of resources. In a case where bank grants the credit to an applicant by pretending it is not risky but in reality it is going to experience the problem of non-payment back, or bank does not grant the credit to an applicant who is not going to have the problem of non-payment back causes that bank makes severe loss.

In traditional credit evaluation method, the main purpose is to grant the credit only to the customers who are good in terms of payment back the loan. Credit application is evaluated by the credit underwriting specialist, if the application is

approved then credit is opened to the applicant. However, in this method, hence the credit decision is given by relying on the subjective judgment of the credit underwriting specialist, customer dissatisfaction is encountered due to the inconsistencies in credit decision system, lack of evaluation of each applicant with the equal and objective factors.

In the new approach in which credit scoring models are employed, the credit applications are evaluated by using decision support system such as scorecards. By using this approach, it is aimed to evaluate the credit applications, whose number is rapidly growing, faster, more effective, and more accurate manner. The speed of the system is associated with objectivity of the credit evaluation and the time spent for decision making. The effectiveness and accuracy depends on the predictive power and consistency of the model that is used in credit evaluation process.

Credit score, which displays the probability that the applicant is defaulted, is the ultimate output of the credit scoring model, which is generated in accordance with the business needs, constructed in this paper. The data obtained from one of a major bank operating in Turkey is splitted into training data with a sample size of %70 percent of the population, and test data with a sample size of %30 percent of the population. Among the credit factors , that are considered by the bank which supplies the data set, the most predictive variables in terms of having statistically significant discrimination power in separation of goods and bads are determined, and by using these selected variable the credit scoring model is constructed on training data. The accuracy of the model is statistically tested by comparing gini coefficients of training and test data. Hence, both data sets have high gini coefficients and these gini coefficients are close to each other, it is concluded that the generated model is valid.

In almost all of the credit scoring applications, Gini Coefficient or Efficiency Index is used to measure the performance of the scorecard. It measures how effectively the scorecard discriminates between Goods and Bads. The reference gini for a scorecard working efficiently is 0.40. The model constructed in this paper has gini value of 0.59 which means the discrimination power of the model is quite good.

Since the generated model is validated and performs well, it is concluded that the model can be used as decision support system in managing the credit risk by banks.

The performance of the model can be improved by taking on different modeling techniques and different predictor variables. Even though the usefulness has not been proved, Reject Inference can be used to make the model more sound which is getting very popular in credit scoring environment. Reject Inference can be shortly described as follows. The probability of good, generated from the model which is built using only accepted applications is calculated for rejected applications since they have no account performance, in modeling terms they have no target variable and this estimate is then adjusted according to the probability of accept which is generated from Accept/Reject model. The principle behind this is that fact that a model which is constructed by only using accepted applications will not be accurate on rejected applications that significantly different from the accept population.

REFERENCES

- Al Amari, A. 2002. The credit evaluation process and the role of credit scoring. Ph.D. Thesis, University College Dublin.
- Altman, Edward I. "Financial ratios, discriminant analysis and the prediction of corporation bankruptcy." *The Journal of Finance*, Volume 23, 1968: 589-609
- Anderson, R. *The Credit Scoring Toolkit*. New York: Oxford University Press, 2007.
- Bailey, M. *Consumer Credit Quality: Underwriting, Scoring, Fraud Prevention and Collections*. Bristol: White Box Publishing, 2004.
- BDDK. "Bankaların İç Denetim ve Risk Yönetim Sistemleri Hakkında Yönetmelik." 2001a.
- Beaver, W. H. Financial ratios As predictors of failure. *Journal of Accounting Research*, 1966: 77-111.
- Breiman, L., J.H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA., 1984.
- Crone, Sven F., and Steven Finlay. "Instance Sampling in Credit Scoring: An Empirical Study of Sample Size and Balancing." *International Journal of Forecasting*, 2011: 224-238.
- Crook, J.N. *Credit Scoring: An Overview*. Working paper series No. 96/13, British Association, Festival of Science. University of Birmingham, The University of Edinburgh, 1996.
- Crook, Jonathan N., Edelman David B., Thomas Lyn C. "Recent Developments in Consumer Credit Risk Assessment." *European Journal of Operational Research*, 2007: 1447-1465.

- Durand, D. *Risk Elements in Consumer Installment Financing*. New York: NBER, 1941.
- Emel, A., M. Oral, A. Reisman, ve R. Yolalan. "A credit scoring approach for the commercial banking sector." *Socio-Economic Planning Sciences*, 2003: 103–123.
- Falkenstein, E. "Credit Scoring for Corporate Debt." In *Credit Ratings: Methodologies, Rationale, and Default Risk*, 169-188. London: Risk Books, 2002.
- Feelders, A. J. "Credit scoring and reject inference with mixture models." *Intelligent Systems in Accounting, Finance and Management*, 2000: 1-8.
- Finlay, Steven. "Credit Scoring for Profitability Objectives." *European Journal of Operational Research*, 2009: 528-537.
- Finlay, Steven. *Credit Scoring, Response Modelling and Insurance Rating*. Palgrave Macmillan, 2010.
- Fischer, R. A. "The Use of Multiple Measurements in Taxonomic Problems." *Ann. Eugenics*, 1936: 179-188.
- Guyon, Isabelle, and Andr e Elisseeff. "An Introduction to Variable and Feature Selection." *Journal of Machine Learning Research*, 2003: 1157-1182.
- Hair, J. F., R. E. Anderson, R. L. Tatham, and W. C. Black. *Multivariate Data Analysis* 5/E. New Jersey: Prentice Hall, 1998.
- Hall, Mark A., and Lloyd A. Smith. *Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper*. New Zealand, 1998.
- Han, J., and M. Kamber. *Data Mining Concepts and Techniques*. San Francisco: Morgan Kaufmann, 2006.
- Hand, D.J. and Henley, W.E. "Statistical Classification Methods in Consumer Credit Scoring." *Royal Statistical Society*, 1997: 523-541
- Heffernan, S. *Modern Banking*. Chichester: John Wiley & Sons Ltd, 2005.
- Henley, W.E and Hand, D.J. A k-nearest neighbor classifier for assessing consumer credit risk. *Statistician*. 1996: 77-95.

- Lewis, E. M. *An Introduction to Credit Scoring*. San Rafael CA: Athena Publishing, 1992.
- Malhotra, R., Malhotra, D. K. "Evaluating consumer loans using Neural Networks." *Omega the International Journal of Management Science*, 2003: 83-96.
- Linder, R., Geier, J. and Kölliker, M.. 'Artificial neural networks, classification trees and regression: Which method for which customer base?'. *Journal of Database Marketing & Customer Strategy Management* 2004: 344–56.
- Patterson, D. W. *Artificial Neural Networks*. New York: Prentice Hall, 1996.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams. "Learning Representations by Back-Propagating Errors." *Nature*, 1986.
- Safavian, S. R., and D. Landgrebe. "A survey of decision tree classifier methodology." *IEEE Trans. Systems Man Cybernetics*, 1991: 660-674.
- Siddiqi, N. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. New Jersey: John Wiley & Sons, Inc., Hoboken, 2006.
- Tan, P. N., M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Boston: Pearson Addison-Wesley, 2006.
- Thomas, L. C., D. Edelman, and J. N. Crook. *Credit Scoring and Its Applications*. Philadelphia: Society for Industrial and Applied Mathematics, 2002.
- West, P. M., Brockett, P. L. and Golden, L. L.. 'A comparative analysis of neural networks and statistical methods for predicting consumer choice'. *Marketing Science* 1997: 370–91.
- Wonderlic, E. F. "An Analysis of Factors in Granting Credit." *Indiana University Bulletins*, 1952: 163-176.