



Developing credit scorecards using logistic regression and classification and regression trees

T Nundhlall

 **orcid.org 0000-0002-5579-1837**

Dissertation accepted in partial fulfilment of the requirements for the degree *Master of Science in Computer Science* at the North-West University

Supervisor: Prof P Pretorius

Graduation: May 2020

Student number: 22288309

Acknowledgements

I would first like to thank my supervisor Professor Philip Pretorius of the Faculty of Natural and Agricultural Sciences at North West University. The door to Prof. Pretorius' office was always open whenever I had an issue or had a question about my research or writing. He consistently allowed this paper to be my own work but directed me in the right direction whenever he thought I needed it.

I would also like to thank the financial institute that allowed me to use their data to complete my thesis.

I would like to thank Gerard Bos for his support, allowing me to use the data and for giving me the opportunity to continue with the research. I would like to thank my colleagues, Augustine Twum, Humbelani Mbedzi, Thobekile Shangase and Thapelo Sekanka for their assistance and guidance during the completion of my thesis.

Finally, I must express my very deep appreciation to my family for providing me with constant support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This achievement would not have been possible without them. Thank you,

Abstract

Financial institutes receive thousands of credit applications daily; thus, consumer credit has become increasingly important in the economy. Credit scoring is the evaluation of the risk associated with granting credit to applicants. Credit scoring is used to predict the probability that a prospective loan applicant or current loan applicant will default or will become delinquent, in other words, it is used to distinguish between good and bad payers. A Scorecard is the tool used in credit scoring, a scorecard is a statistical model which considers the correlation between all different characteristics of an historic behaviour of the applicant and tries to predict the applicant's future behaviour. Various data mining techniques are used to build a scorecard.

Before developing the scorecard, the data needs to be extracted and cleaned. A Masterfile analysis is then conducted to determine the Good/Bad definition; to achieve this the performance window is used to monitor accounts opened in that period to determine if they went bad or not. The sample window is the period used to develop the scorecard. The roll rates analysis is used to confirm the definition. This is done by comparing the worst delinquency status in a specific month x to the delinquency status in the next month and by then calculating a percentage of accounts that maintained their delinquency status, "rolled forward" into the next delinquency status or got better. Once the definition is confirmed, development of the scorecard may begin.

Logistic regression is the most commonly used technique in the market for the development of a scorecard. In a logistic regression the dependent variable makes the assumption that the event of interest has occurred or has not occurred. When building a credit scoring model using a logistic regression model; outliers are not present because continuous predictors are converted to uniform scores, no correlation of more than 0.5 may be present between predictors. The aim of logistic regression is to find the best fitting model to describe the relationship between the dependent variable and a set of independent variables, the outcome variable of this model is binary.

Although logistic regression is the most commonly used statistical technique in building scorecards, other techniques can also be used, such as Classification and Regression Trees (CART), it is a machine learning technique which is non-parametric and is generally used in predictive modelling. It is a step by step process which constructs a decision tree by splitting or not splitting each node on a tree into two daughter nodes. CART can discover complex interactions between predictors which might be impossible when using traditional techniques.

CART uses binary recursive splitting, where the dependent variable is categorical and the “class” is where the dependent variable falls into, is classified by the tree.

Scorecards were developed using the logistic regression and CART methods respectively. The logistic regression method had performed better than the CART method. When using the methodology proposed in the following research the logistic regression model performed better; the logistic regression produced a stronger Gini, selected variables that were more stable over time and selected variables that had no correlation between them.

Keywords: Credit Scoring, Logistic Regression, Classification and Regression and Trees, model

Table of Contents

CHAPTER 1: INTRODUCTION.....	1
1.1 Introduction	1
1.2 Problem statement.....	2
1.3 Objectives of the study.....	3
1.3.1 Primary objectives	3
1.3.2 Theoretical objectives	3
1.3.3 Empirical objectives.....	3
1.4 Research design and methodology.....	3
1.4.1 Literature review	4
1.4.2 Empirical study.....	10
1.4.3 Sampling Methodology	10
1.4.4 Development sample	11
1.4.5 Sample Period.....	13
1.4.6 Validation and Stability.....	13
1.4.7 Recent Sample – For Stability Tests	14
1.5 Statistical analysis.....	14
1.6 Ethical considerations	15
1.7 Chapter classification	15
1.7.1 Chapter Classification Flow Chart.....	15
CHAPTER 2: DEVELOPMENT AND IMPLEMENTATION PROCESS OF AN ORIGINATION SCORECARD	17
2.1 Stage 1: Data Exploration.....	18
2.1.1 Initial characteristic analysis: Known Good Bad Model	18
2.2 Preliminary Scorecard: Known Good Bad model (KGB)	19
2.3 Reject Inference	19
2.4 Final Scorecard: All Good Bad model (AGB)	20

2.4.1 Assignment of Points	20
2.5 Validation.....	20
2.5.1 Scorecard stability report.....	21
2.5.2 Variable analysis report	22
2.6 Post-Implementation.....	22
2.7 Summary	24
CHAPTER 3: DATA SOURCES AND PERFORMANCE DEFINITIONS.....	25
3.1 Introduction	25
3.2 Data sources.....	26
3.2.1 Application data source	26
3.2.2 Account performance data source	26
3.3 Performance definition	27
3.3.1 Good/bad/indeterminate/exclusion definition	27
3.3.2 Roll rate analysis.....	28
3.4 Summary	36
CHAPTER 4: LOGISTIC REGRESSION.....	38
4.1 Introduction	38
4.2 Assumptions of a logistic regression.....	38
4.3 Logistic regression results	39
4.4 The Gini coefficient.....	40
4.5 Known good bad (KGB) model	40
4.5.1 KGB variables with Gini	40
4.5.2 KGB correlation matrix.....	41
4.5.3 KGB validation graph	42
4.5.4 Variance inflation factors	43
4.6 Accept reject (AR) model.....	44
4.6.1 AR variables with Gini.....	44
4.6.2 AR correlation matrix.....	44

4.6.3 Variance inflation factors	45
4.7 Reject inference	45
4.7.1 Reject inference process	46
4.7.2 Reject inference checks.....	48
4.8 All Good Bad (AGB) model.....	50
4.8.1 AGB variables with Gini	50
4.8.2 AGB correlation matrix.....	51
4.8.3 AGB validation graph	51
4.8.4 Variance inflation factors	52
4.9 Weights of Evidence trend.....	54
4.9.1 Weights of evidence graph.....	55
4.10 Stability over time analysis	66
4.11 Summary.....	71
CHAPTER 5: CLASSIFICATION AND REGRESSION TREES.....	72
5.1 Introduction	72
5.2 Classification and decision factors	72
5.3 Binary recursive partitioning.....	72
5.4 Pruning	73
5.5 Classification and regression tree	74
5.5.1 Variables with Gini.....	75
5.5.2 Importance of the variables.....	76
5.5.3 Cumulative lift.....	78
5.5.4 Classification tables.....	79
5.5.5 Fit statistics	80
5.5.6 The assessment plot	82
5.5.7 The leaf statistics plot.....	83
5.6 Summary	85

CHAPTER 6: SUMMARY, CONCLUSION AND RECOMMENDATIONS	86
6.1 Introduction	86
6.2 Chapter 1: Summary	86
6.2.1 Introduction	86
6.2.2 Problem statement	86
6.2.3 Objectives of the study.....	86
6.2.3.1 Primary objectives	86
6.2.3.2 Theoretical objectives	86
6.2.3.3 Empirical objectives	87
6.2.4 Research design and methodology	87
6.2.4.1 Empirical study.....	87
6.2.4.2 Statistical analysis	87
6.2.5 Ethical considerations	87
6.3 Chapter 2: Summary	88
6.3.1 Introduction	88
6.3.2 Stage 1: Data Exploration	88
6.3.3 Preliminary Scorecard: Known Good Bad Model (KGB)	88
6.3.4 Reject Inference.....	88
6.3.5 Final Scorecard: All Good Bad model (AGB)	88
6.3.6 Validation	88
6.3.7 Post-Implementation	89
6.4 Chapter 4: Summary	89
6.4.1 Introduction	89
6.4.2 Assumptions of a logistic regression	89
6.4.3 Logistic regression results.....	89
6.4.4 The Gini coefficient.....	89
6.4.5 Known Good Bad (KGB) model	89
6.4.6 Accept Reject (AR) model.....	90

- 6.4.7 Reject Inference90
- 6.4.8 All Good Bad (AGB) model90
- 6.4.9 Weights of Evidence trend90
- 6.4.10 Stability over time analysis90
- 6.5 Chapter 5: Summary90
 - 6.5.1 Introduction90
 - 6.5.2 Classification and decision problems90
 - 6.5.3 Binary and recursive partitioning91
 - 6.5.4 Pruning91
 - 6.5.5 Classification and regression tree91
- 6.6 Results91
- 6.7 Conclusion and recommendations92
- BIBLIOGRAPHY94

List of Figures

Figure 1 1: Population flow - Development sample – prior to exclusions of policy rules declines ...	11
Figure 1 2: Population flow - Development sample – after exclusions of policy rules declines	12
Figure 1 3: Observation and Outcome period	13
Figure 1 4: Chapter classification flow chart	15
Figure 2. 1: Scorecard development process	17
Figure 2. 2: Validation chart.....	21
Figure 3. 1: Data extraction periods.....	26
Figure 3. 2: Roll rates analysis 12 months - Overdrafts	29
Figure 3. 3: Roll rates analysis 12 months – Revolving Loans	29
Figure 3. 4: Roll rate analysis 18-month definition – Revolving Loans	30
Figure 3. 5: Roll rates analysis 18-month definition - Overdrafts.....	31
Figure 3. 6: Marginal bad rate change – Overdrafts.....	32
Figure 3. 7: Marginal bad rate change – Revolving Loans	33
Figure 3. 8: Population flow - Development sample.....	34
Figure 3. 9: Observation and Outcome period	35
Figure 3. 10: Population flow Development sample - Existing to Bank, Thick	36
Figure 4. 1: Known Good Bad model validation graph.....	42
Figure 4. 2: Reject inference process	46
Figure 4. 3: Log odds chart - Reject inference results.....	47
Figure 4. 4: All Good Bad model validation graph.....	52
Figure 4. 5: WoE graph Var 1	55
Figure 4. 6: WoE graph Var 2	56
Figure 4. 7: WoE graph Var 3	57
Figure 4. 8: WoE graph Var 4	58
Figure 4. 9: WoE graph Var 5	59
Figure 4. 10: WoE graph Var 6	60
Figure 4. 11: WoE graph Var 7	61
Figure 4. 12: WoE graph Var 8	62
Figure 4. 13: WoE graph Var 9	63
Figure 4. 14: WoE graph Var 10	64
Figure 4. 15: WoE graph Var 11	65

Figure 4. 16: Stability over time graph Var 1	66
Figure 4. 17: Stability over time graph Var 2	66
Figure 4. 18: Stability over time graph Var 3	67
Figure 4. 19: Stability over time graph Var 4	67
Figure 4. 20: Stability over time graph Var 5	68
Figure 4. 21: Stability over time graph Var 6	68
Figure 4. 22: Stability over time graph Var 7	69
Figure 4. 23: Stability over time graph Var 8	69
Figure 4. 24: Stability over time graph Var 9	70
Figure 4. 25: Stability over time graph Var 10	70
Figure 4. 26: Stability over time graph Var 11	71
Figure 5. 1: Cumulative lift of Classification and Regression Tree model.....	78
Figure 5. 2: Assessment plot of Classification and Regression Tree model	82
Figure 5. 3: The leaf plot of the Classification and Regression Tree model.....	83

List of Tables

Table 3. 1 Good/Bad definition – Overdrafts.....	27
Table 3. 2: Good/Bad definition – Revolving Loans	28
Table 3. 3: Roll rates analysis – Revolving Loans.....	30
Table 3. 4: Roll rates analysis - Overdrafts	32
Table 4. 2: List of Known Good Bad model variables and Gini’s.....	41
Table 4. 3: Known Good Bad model correlation matrix.....	41
Table 4. 4: List of Known Good Bad variables and VIF values.....	43
Table 4. 5: List of Accept Reject model variables and Gini’s.....	44
Table 4. 6: Accept Reject model correlation matrix.....	44
Table 4. 7: List of Accept Reject model variables and VIF values.....	45
Table 4. 8: Bad rate adjustment.....	48
Table 4. 9: Bad rate improvement	49
Table 4. 10: Gini improvement.....	49
Table 4. 12: All Good Bad model correlation matrix.....	51
Table 5. 1: Classification and Regression Tree model variables.....	76
Table 5. 2: Classification and Regression Tree model variable importance list.....	77
Table 5. 3: Classification table for training data set in Classification and Regression Tree model..	79
Table 5. 4: Classification table for validation data set in Classification and Regression Tree model	79
Table 5. 5: Fit statistics of Classification and Regression Tree model.....	80

Abbreviations

Abbreviation	Meaning
AFF	Affordability
AGB model	All Good Bad model
AR model	Accept Reject model
ASE	Average Squared Error
BRS	Behavioural Risk Score
CART	Classification and Regression Trees
ETB	Existing to Bank
F12	First 12 Months
F18	First 18 Months
FR	Risk Rating
GB	Good/Bad
IV	Information Value
KGB	Known Good Bad Model
Log/Ln	Logarithmic function
MBG	Maintenance Bureau Gateway
NTB	New to Bank
NTU	Not Taken Up
OD	Overdraft
OoT	Out of Time
PD	Probability of Default
PIT	Point in Time
PP	Payment Profile
Rej	Rejected
RL	Revolving Loan
S18	Second 18 months
TTC	Through the Cycle
TU	Taken Up
Var	Variable
VIF	Variance Inflation Factor
WoE	Weights of Evidence

CHAPTER 1: INTRODUCTION

In this chapter the issues mentioned are discussed at a high level and detailed in the later chapters.

1.1 Introduction

Credit scoring is the statistical process used in the financial world to estimate the likelihood of an applicant repaying any credit approved to them. Existing information about the applicant is used; such information includes the applicant's bureau, demographic and behavioural data which is used to develop a numerical score for each applicant.

Credit scoring is used to determine an applicant's credit worthiness by using the applicant's history on the performance of his/her accounts to predict the performance of the applicant in the future. When assessing an applicant for credit the following is taken into account; the applicant's character, the applicant's ability to pay, the applicant's capital, the collateral provided by the applicant and the condition of the applicant's business or finances. Credit scoring is applied throughout the applicant's life cycle.

Scorecards are the tools used in credit scoring, a scorecard is a statistical model which takes into account the correlation between different aspects of an applicant's past behaviour aiming to try and predict the future behaviour of the applicant. There are various data mining techniques that can be used to build a scorecard.

Data mining is a logical approach to find principal patterns, trends and relationships concealed in the data. Data mining can be characterised into two categories, methodologies and technologies. Methodologies involve data visualisation, machine learning, statistical techniques and deductive database. The applications that use these methodologies can be summarized as classification, prediction, clustering, summarization, dependency modelling, linkage analysis and sequential analysis. The technology part of data mining consists of techniques such as statistical methods, decision trees, genetic algorithms, neural networks and non-parametric methods (Lee et al., 2006:1114-115). Classification methods in which observations can be assigned to one of numerous disjoint groups are vital in business decision making because of their vast applications in decision support, financial forecasting, fraud detection, marketing strategy, process control and other related fields (Lee et al., 2006:1114).

This study will compare two data mining modelling techniques to build a scorecard, namely, Classification and Regression Trees (CART) and Logistic Regression.

Logistic regression is the most commonly used technique in building credit scoring models (scorecards). Logistic regression is used to find the best fitting model that is the most parsimonious and interpretable to describe the relationship between a dependent (outcome) variable and a set of independent variables. In logistic regression the outcome variable is dichotomous (0/1 outcome).

CART analysis is a decision tree building technique which can help identify the most significant variables. Decision trees are built on historical data with pre-assigned classes for all observations. A crucial aspect to CART is recursive partitioning. CART often reveals complex interactions between predictors which may be complicated or impossible using traditional multivariate methods, for example the value of a variable (e.g. time) may significantly influence the importance of another variable (e.g. age). These interactions are not the easiest to model as the number interactions and variables begin to increase. CART is also robust to outliers. The structure of the decision trees remains unchanged with respect to monotone transformations of the independent variables.

The aim of this study is to compare the results of the two data mining modelling techniques used to build a scorecard.

1.2 Problem statement

The procedure of credit scoring is of high importance in financial institutes due to the fact that they need to separate good applicants from bad applicants in terms of their creditworthiness. Methods normally used for credit scoring are founded on statistical pattern recognition techniques. These refined methods challenge the perception that regards credit scoring as being a simple data mining process. Statistical pattern recognition is a rising research area, which suggests that most existing statistical models applied to credit risk are being compared with each other to see which method will predict the best possible results.

The key problem of any financial institute is to distinguish between good and bad applicants prior to granting credit; which can be done by using credit scoring modelling techniques. The

goal of this study is to compare the results of two credit scoring techniques; namely Classification and Regression Trees (CART) and Logistic Regression; and to elaborate on which method seems to be the most efficient and effective. By doing this comparison it will help to understand the differences in the modelling techniques and how they work. We will be able to find out which technique predicts the outcome more accurately, thus increasing sales in the financial institute and lessening the risk of lending to applicant's that would default over time. The two scorecards will be built using the exact same dataset and time periods. The results will then be compared.

1.3 Objectives of the study

The following objectives have been formulated for the study:

1.3.1 Primary objectives

The primary objective of this study is to develop scorecards using two different statistical modelling techniques and comparing the results of the two techniques' as well as the performance, advantages and disadvantages.

1.3.2 Theoretical objectives

In order to achieve the primary objective, the following theoretical objectives are formulated for the study: (a) to understand what credit scoring entails and (b) to understand the two modelling techniques being compared.

1.3.3 Empirical objectives

In accordance with the primary objective of the study, the following empirical objectives are formulated: to develop and compare scorecards using two different modelling techniques and comparing results. The aim is to find an effective modelling technique that will have a stronger and more accurate predictive power.

1.4 Research design and methodology

The study will comprise a literature review and empirical study. The study will be one of a quantitative nature, which will be used for the empirical part of the study.

1.4.1 Literature review

The following section will be based on literature reviews based on credit scoring, logistic regression and classification and regression trees.

Credit Scoring

Consumer credit has become increasingly important in the economy. Credit scoring is used to predict the probability that a prospective loan applicant or current loan applicant will default or become delinquent, i.e. distinguish between the good and bad payers. According to Thomas (Thomas et al., 2002:1) credit scoring is a series of decisioning models and their methodology that help in granting credit. According to Mester (Mester Loretta, 1997:4) credit scoring is used to assess the risk of a loan application; this is done by using historic data and statistical methods, credit scoring tries to identify the effects of different characteristics on the loan application. A score is produced that a financial institute uses to rank its loan applicants in terms of risk. Credit scoring has gained attention since the credit industry can benefit from improving cash flow, ensuring credit collections and reducing potential risks. There are so many papers that used intelligent and statistical techniques since the 1930s (Siami and Hajimohammadi, 2013:119). In the 1930s, numerical scorecards were first used by mail-order companies. Since then; the use of data mining techniques has improved in this research area and has become the dominant area in the field of credit scoring.

When assessing credit, we can roughly summarize the different kinds of scoring as follows (Siami and Hajimohammadi, 2013:119-120):

- Application credit scoring: is the assessment of the credit worthiness of prospective applicants. "It quantifies the default associated with credit requests by questions in the application form, e.g. salary, number of dependents, etc."
- Behavioural scoring: is similar to application scoring, except it refers to existing applicants. "The decision on how the lender needs to deal with the borrower is in this area. Behavioural scoring models use the applicant's historical data, e.g. account activity, account balance and age of the account etc."
- Collection scoring: "is used to divide the applicants with different levels of insolvency into groups, separating those who require more crucial actions from those who don't need to be attended to immediately. The models are distinguished according to the degree of delinquency and allow better management of delinquent applicants, from the first signs of delinquency to subsequent phases and debt write-off."

- Fraud detection: “models rank the applicants according to the relative likelihood that an application may be fraudulent.”

The development of a scorecard is the first step in credit scoring. To help with managing large amounts of data, the data is classified based on where it was sources i.e. internal or external. Classification model/methodologies are used to build credit scoring models that use predictors to estimate the probability of default, two commonly used methods are the logistic regression and classification or decision trees, (Barbon and Vidal, 2019).

Risk scorecards thus offer a powerful and empirical solution to business needs. Many industries use risk scorecards to predict delinquency non-payment, fraud, insurance claims and the recovery of amounts owed for accounts in collections. Scoring methods offer an objective and consistent approach to assess risk if the system overrides are kept as low as possible.

According to (Anderson, 2007), credit scoring can be defined as a model that helps with the decision-making process when scoring an application applying for credit. It is subjective and through technological advances the need for credit rating decisions to be objective and reliable are made possible. Using a credit score allows the financial institute (lender) to treat each applicant objectively therefore the same standards apply to every applicant. The underlying techniques used in credit scoring help in deciding who will be granted credit, how much the credit granted should be and what business strategies will improve the profitability of the credit granters, (Thomas et al., 2002). The general idea of credit scoring is to compare the features or characteristics of prospective applicants with current applicants, whose loans have already been paid back. If a prospective applicant’s characteristics are sufficiently similar to current applicants who have been approved credit and defaulted, then the prospective applicant will usually be declined. Likewise, if a prospective applicant’s features or characteristics are sufficiently similar to current applicants who have no defaulted then the application will generally be granted/approved.

Credit scoring models are the most successful models used in financial institutions. In a credit scoring model, according to Abdou and Pointon, analysts usually use their historical experience with debtors to derive a quantitative model for the segregation of acceptable and unacceptable credit applications (Abdou and Pointon, 2011:3). Thus, when using a scoring system, the credit application is a self-operating process and is constantly applied to all credit decisions, therefore

allowing financial institutes to assess creditworthiness more rapidly. Regardless of the criticisms of a credit scoring model there are benefits as well thus it can still be regarded as one of the most successful models in the financial world.

According to Siddiqi, the most powerful and practical solution to solving business needs with regard to credit scoring is the risk scorecard, (Siddiq, 2006).

Benefits of credit scoring

Credit scoring requires less information to make a decision, because credit scoring models have been estimated to include only those variables, which are statistically and/or significantly correlated with repayment performance. Credit scoring models attempt to correct the bias that would result from considering the repayment histories of only accepted applications and not all applications. This is done by assuming how rejected applications would have performed if they had been accepted (Abdou and Pointon, 2011:4). Credit scoring models consider the characteristics of good as well as bad payers. Credit scoring models are built on much larger samples than a loan analyst can remember. Credit scoring models can be seen to include explicitly only legally acceptable variables whereas it is not so easy to ensure that such variables are ignored by a loan analyst. Credit scoring models demonstrate the correlation between the variables included and repayment behaviour. A credit scoring model includes a large number of a customer's characteristics simultaneously, including their interactions, while a loan analyst's mind cannot arguably do this, for the task is too challenging and complex. An additional essential benefit of credit scoring is that the same data can be analysed easily and clearly by different credit analysts or statisticians and give the same weights (Abdou and Pointon, 2011:3-5). A few other benefits of credit scoring are that it provides support in the decision making process which can lead to fewer errors being made. The modelling is based on factual data where the interrelation between the variables are considered and analysed. The cut-off score may also be adjusted according to environmental factors that affect the financial institute.

Criticisms of credit scoring

Credit scores use any characteristic of a customer in spite of whether a clear link with a likely repayment can be justified. Also, sometimes economic factors are not included. In addition, using credit scoring models, sometimes customers may have the characteristics, which make them more similar to bad than good payers, but may have these entirely by chance (a

misclassification problem). Statistically a credit scoring model is “incomplete”, for it leaves out some variables, which taken with the others, might predict that the customer will repay. But unless a credit scoring model has every possible variable in it, normally it will misclassify some people (Abdou and Pointon, 2011:5).

Thus, a customer who applies for credit is evaluated in a credit scoring system by summing up points that are received on the different application features to have a total score. Based on how the system is designed this total score may be treated in various ways. In a single cut-off method, the applicant’s score is compared to a specific cut-off score. If this score exceeds the cut-off score, then credit is granted if not then credit is denied to the applicant. Highly developed credit scoring systems are based on a two-stage process. For example, the applicant’s total score may be compared to two cut-off points. If the score exceeds the higher cut-off, credit is granted, if the total score falls below the lower cut-off, credit is rejected. If the total score is between the two cut-off points, the analyst re-evaluates the applicant based on the specific requirements, or alternatively, historic credit data is collected, then scored, the points are then added to the total score, the information is obtained from the completed application form. As a result of this approach, credit is granted if the new score is above the new cut-off, if not, the credit is declined, (Capon, 1982).

Credit scoring models are only as good as the specification that is given; another limitation is that the data is historical. Either the variables or the weights or both, are assumed to be stable over time, making the model less accurate, unless it is updated regularly. To reduce this problem, it has been recommended that financial institutes keep records of the type I and Type II errors and a model that has been updated or a new model should be applied to accommodate the new changes.

Logistic regression

Logistic regression is the most used technique in the market for the development of credit scoring models. Logistic regression competes with discriminant analysis as a technique for analysing discrete response variables. Most statisticians of today feel that logistic regression is more versatile and is more appropriate for most situations than discriminant analysis because logistic regression does not assume that the independent variables are normally distributed whereas discriminant analysis assumes this. The essential mathematical model that underlies logistic regression is the logit; which is known as the natural logarithm of an odds ratio. The

aim of logistic regression is to find the best fitting model to describe the relationship between the dependent variable and a set of independent variables; quite often the outcome variable is discrete, taking on two or more possible values. The independent variables are also called covariates. Logistic regression is a fairly robust, flexible and easily used method which can derive important interpretations. What makes a logistic regression different is the outcome variable, which is dichotomous. This difference can be seen in both the model and its assumptions compared to other regression methods such as a linear regression model. Once this difference is accounted for, the methods employed in an analysis using logistic regression follow, more or less, the same general principles used in linear regression (Hosmer Jr et al., 2013:1).

The goal of logistic regression is to accurately predict the dichotomous outcome for individual cases using the most parsimonious model. Logistic regression tries to find a relationship between a dependent variable (Y) with one or more independent variables (x_i) which are categorical or continuous and related to the occurrence of an event (Callejon-Ferre et al., 2019). To achieve this goal, the model includes all predictor variables that are useful in predicting the response variable. A number of different options are available during model creation. Variables can be entered into the model in the order specified by the analyst/researcher or logistic regression can test the fit of the model after each coefficient that is added or deleted, this is called a stepwise regression. Stepwise regression is normally used in the exploratory phase of research, but it is not advisable to use for theory testing. Theory testing is the testing of “apriori theories” or hypotheses of the relationships between variables. Exploratory testing makes no “apriori assumptions” regarding the relationships between the variables, thus the goal is to discover relationships.

The explicit form of the logistic regression model for one variable x hat will be used is:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1.1)$$

There are quite a few parts involved in the evaluation of a logistic regression model. First, the final model (the relationship between all of the independent variables and dependent variable) needs to be assessed. Second, the significance of each of the independent variables needs to

be assessed. Third, predictive accuracy or discriminating ability of the model has to be evaluated. Finally, the model needs to be validated.

Classification and Regression Trees (CART)

Although logistic regression is the most commonly used statistical technique in building credit scoring models, other techniques can also be used such as CART. CART is non-parametric and was introduced by Breiman and his colleagues in 1984. The main idea of CART is recursive partitioning of the data into smaller and smaller strata to help improve the fit as best as possible. In recent research CART has been found to be rather effective for creating decision rules which perform just as well if not even better than rules developed using traditional methods CART is often able to discover complex interactions between predictors which might be impossible using traditional multivariate techniques.

To build decisions trees CART uses a learning set, this is a set of historical data with pre-assigned classes for all observations. Recursive partitioning is the key to the nonparametric statistical method of CART (Anonymous, 2011:223). It is a step by step process which constructs a decision tree by splitting or not splitting each node on a tree into two daughter nodes. The CART methodology has an attractive feature which is; because the algorithm asks a series of hierarchical questions it makes it easier and simpler to understand and interpret the results. The root node is the unique starting point of the classification tree and consists of the entire learning set \mathcal{L} . A node is a subset of the set of variables, which can be terminal or non-terminal. A non-terminal node also known as the parent node is a node that splits into two daughter nodes, this is also known as a binary split. The binary split is determined by a condition on the value of a single variable, the condition is either satisfied or not satisfied by the observed value of that variable. All the observations in \mathcal{L} which have reached a particular parent node and satisfies the condition for that variable, drop down to one of the daughter nodes; the outstanding observations at that parent node that don't satisfy the condition drop down to the other daughter node. Nodes that do not split are called terminal nodes and are assigned a class label; each observation in \mathcal{L} ends up in one of the terminal nodes. An observation of an unknown class that is "dropped down" the tree and ends up at a terminal node is assigned the class equivalent to the class label attached to that node. It is possible to have more than one terminal node with the same class label.

To produce a tree-structured model using recursive binary partitioning, CART determines the best split of the learning set \mathcal{L} to start with and thereafter the best splits of its subsets on the basis of various issues such as identifying which variable should be used to create the split, and determining the precise rule for the split, determining when a node of the tree is a terminal node, and assigning a predicted class to each terminal node (Anonymous, 2011:224). According to Dupuy (Dupuy et al., 2019:2), the splitting method used on each node until it has reached the stopping criteria or for as long as the node is not “pure”. The splits are identified by the values of the inputs. Assigning predicted classes to the terminal nodes is simple, and determining how to make the splits, however determining the right-sized tree is not so easy. There are methods used for “growing” a fully expanded tree and obtaining a tree of optimum size.

1.4.2 Empirical study

The scorecard that will be built will be for Overdrafts and Revolving Loans facility for a financial institute. The model that will be built will be for existing customers applying for an Overdraft or Revolving Loan facility or for a limit increase on a current Overdraft or Revolving Loan facility. The scorecard will determine the risk at application stage of applicants applying for an Overdraft or Revolving Loan and/or limit increase by combining demographic information captured at time of application with internal company Behavioural Risk Scoring information (which will be referred to as BRS information hereafter) and Maintenance Bureau Gateway data or more commonly known as bureau data (which will be referred to as MBG aggregations hereafter) where available time of application.

Data was extracted from relevant tables for application, affordability and performance data. The necessary joins were made using a unique key to join the relevant tables. Duplicates were removed from the final table/sample created.

1.4.3 Sampling Methodology

The sample generally contains $\pm 10\%$ of Good, Bad, Indeterminate and Reject applications and should consist of at least 5000 records for each. In instances where 5000 records are not available then all available records should be included in the sample.

Sampling is then conducted in SAS using the SurveySelect procedure with the GBI definition (indicating whether the record is Good, Bad, Indeterminate, Reject or Not Taken Up) specified as the Strata and then using Simple Random Sampling for each definition in the Strata. Specifying the Strata statement within the SAS procedure partitions the dataset into non-

overlapping groups defined by the GBI definition variable. Independent samples are then selected from these strata. Simple Random Sampling selects units with equal probability and without replacement, which means that a unit cannot be selected more than once from each stratum.

1.4.4 Development sample

The development sample breakdown prior to the exclusion of policy rule declines is provided below. Note that the sample counts are indicated in the brackets

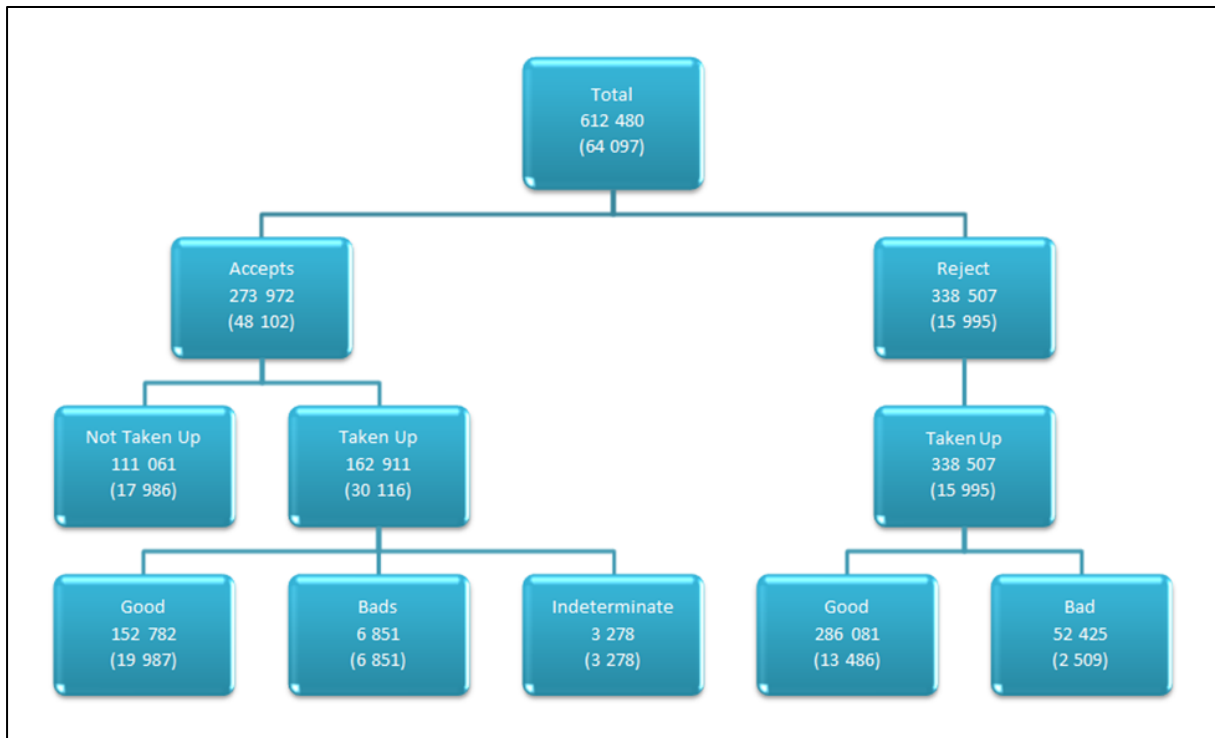


Figure 1 1: Population flow - Development sample – prior to exclusions of policy rules declines

From figure 1.1 it can be seen that a total population of 612 480 with 273 972 accepted customers and 338 507 rejected customers, the acceptance rate was 44.73%. From the accepted customers 111 061 customers had not taken up the product and 162 911 customers had taken up the product. From the accepted customers that had taken up the product; 152 782 were good customers, 6851 were bad customers and 3278 customers were indeterminate customers. From the rejected customers 338 507 had taken up the product and 286 081 of these customers were good and 52 425 were bad customers.

Development Sample – subsequent to policy rule declines

The development sample breakdown provided below. Note that the sample counts are indicated in the brackets.

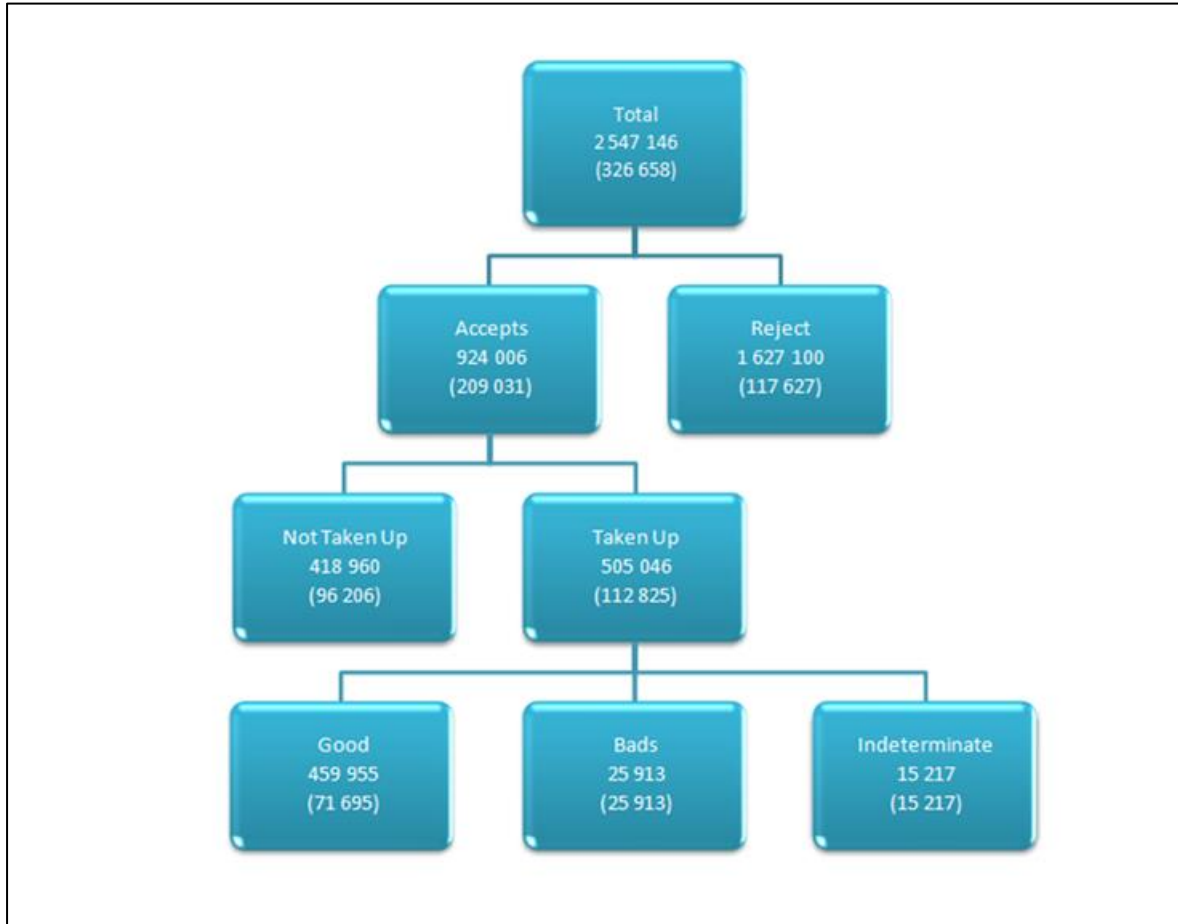


Figure 1 2: Population flow - Development sample – after exclusions of policy rules declines

Segmentation resulted in the development of 5 scorecards for the Overdraft and Revolving Loan portfolio. This study will focus on one segment, namely; existing to bank thick clients. Existing to bank customers are customers who are banking with the financial institute for more than six months and thick refers to the amount of bureau information that the institute has on the customer.

1.4.5 Sample Period

The sample period chosen was from 1 May 2013 to 31 October 2014. This was done based on the fact that portfolio takes 18 months to mature, maturity referring the time it takes for the outstanding loans to be repaid and also keeping in mind that we would have to perform an Out of Time (OoT) validation can be done later on. The maturity period is used for the outcome period to observe the applicant's behaviour.

Figure 1.3 below shows the following:

- The sample window or observation period of 1 May 2013 to 31 October 2014;
- The outcome period, 1 November 2014 to 30 April 2016;
- Maturity of 18 months.

Note that the scorecard development was done on 80% of the sample with 20% used for the holdout validation.

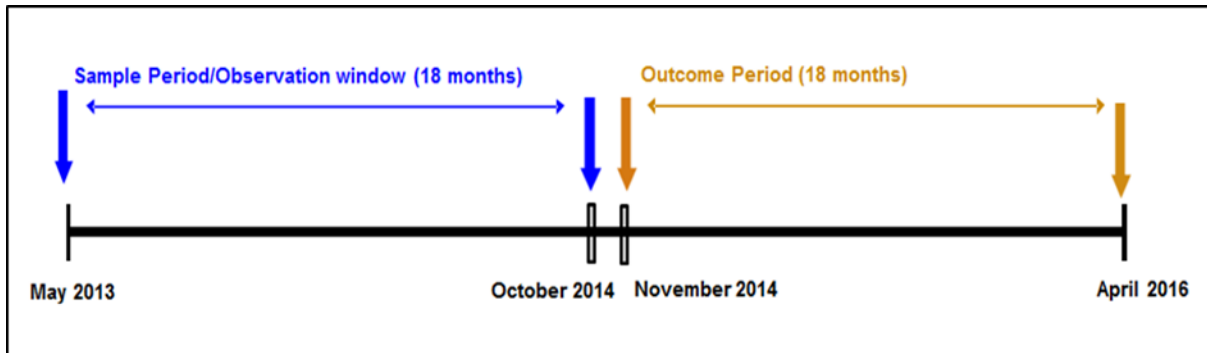


Figure 1 3: Observation and Outcome period

A sample/observation period is the time period from which we sample when building a scorecard. Information on applicants is collected during a certain time period in the past. This information is used to build the scorecard. The outcome/performance period is the time period directly after the sample period; this is the time period allowed to observe the behaviour of the applicants that were accepted for credit during the sample period.

1.4.6 Validation and Stability

There are a variety of ways to divide the development (sample the scorecard is developed on) and validation ("hold") datasets. Usually, 70% to 80% of the sample is used to develop a

scorecard; the remaining 20% to 30% is put aside and subsequently used to independently test or validate the scorecard. Instances where the sample size is too small, the scorecard can be developed by using 100% of the sample size and when validating the scorecard select samples of 50% - 80% that were randomly selected. Validation is done to verify if the model developed is appropriate to the subject population, and to make sure that the model has not been over fitted.

A scorecard can be very predictive at the time of development however variables included in the model could show significant shifts in their distributions over time. We therefore try to include stable variables at the development stage, even if it means giving way some of the strength of the scorecard. Therefore, stability over time analysis is conducted for each variable. If the scorecard is not stable from a volume perspective the expected performance results could also be out of line.

Note that the scorecard development was done on 80% of the sample with 20% used for the hold-out validation.

1.4.7 Recent Sample – For Stability Tests

A recent sample test is done to test if the current population still scores similar to the development sample because the scorecard is normally developed on data that is 2 to 3 years old. The Recent sample was taken from the time period 1 July 2016 to 29 September 2016.

1.5 Statistical analysis

Logistic Regression Approach

The scorecard development using logistic regression will be performed using Paragon's scorecard modelling tool, Modeller, under license from Paragon. The modelling approach used is Logistic Weights of Evidence (WoE) regression modelling with stepwise elimination.

CART Approach

The scorecard development using CART will be performed using SAS Enterprise Miner, under license from SAS.

1.6 Ethical considerations

All the information that is used regarding the description of the variables will be kept discrete and thus renamed due to the sensitivity of the data used. Authorisation has been given to use actual variables provided they are renamed and used in an appropriate manner.

1.7 Chapter classification

This study will comprise the following chapters below, an explanation of the flow chart will follow subsequently.

1.7.1 Chapter Classification Flow Chart

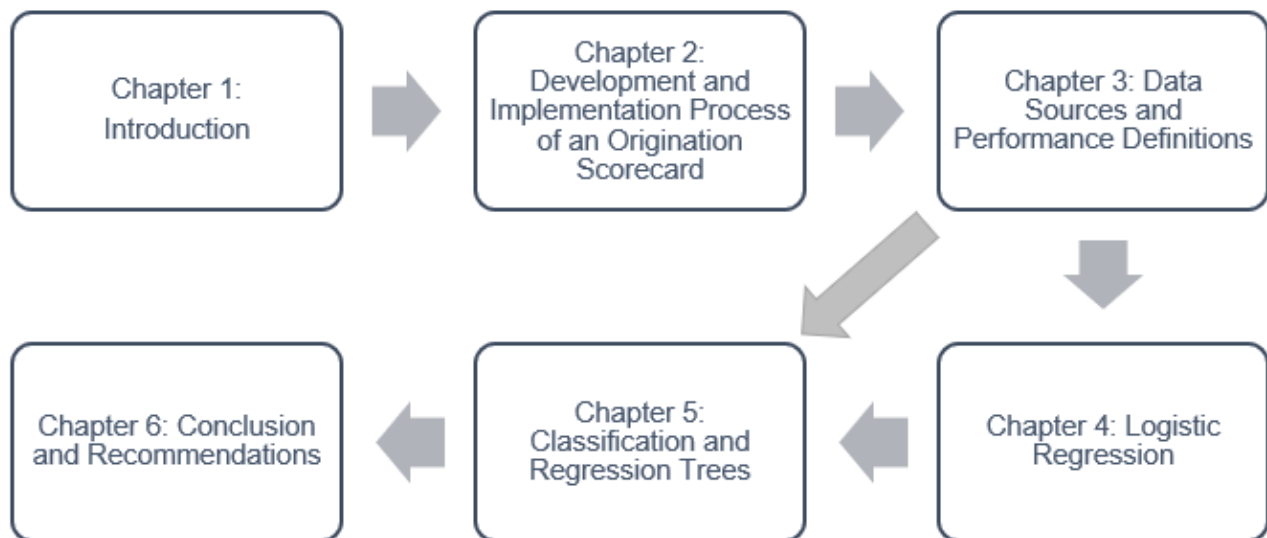


Figure 1 4: Chapter classification flow chart

Chapter 1: Introduction

This chapter has described what the study is about and what the rest of the study will entail. An explanation on how the following chapters relate to form an integrated whole will be given.

Chapter 2: Development and Implementation Process of an Origination Scorecard

This section will describe and explain what credit scoring is, and the credit scoring process.

Chapter 3: Data Source and Performance Definition

This part of the paper will explain the data part of the study. How the data was sampled, and the techniques used in the study to develop the scorecard.

Chapter 4: Logistic Regression

This section of the paper will describe and explain what logistic regression is and what it entails and the results that were produced from the logistic regression.

Chapter 5: Classification and Regression Trees (CART)

This chapter will describe and explain what CART is and what it entails and the results that were produced from the CART method.

Chapter 6: Conclusions and Recommendations

This chapter will explain how the conclusion was reached and which modelling technique showed a better improvement.

CHAPTER 2: DEVELOPMENT AND IMPLEMENTATION PROCESS OF AN ORIGATION SCORECARD

This chapter will cover the process of how an application origination scorecard is built. the following will be discussed in the following chapter, the main steps of building an origination scorecard from data exploration, characteristic selection, to building the Known Good Bad model (KGB), to conducting rejecting inference and finally building the All Good Bad model (AGB), along with the validation and reporting on the stability of the scorecard and analysis of the variables.

Credit scoring is used to predict if a prospective loan applicant would be approved or declined for a credit loan, that the customer has applied for, i.e. to differentiate between the good and bad payers.

There are primary stages in the development and implementation process of a scorecard. The primary stages are as follows:

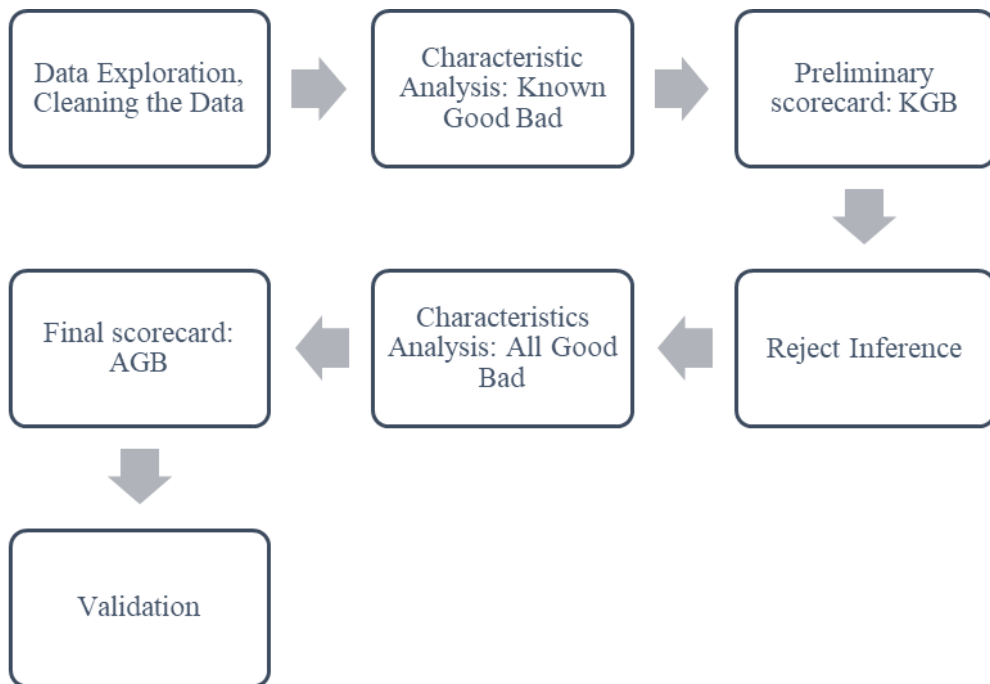


Figure 2. 1: Scorecard development process

2.1 Stage 1: Data Exploration

A reliable and successful scorecard is dependent on the quality of the data used to build the scorecard. Obtaining insights during the development of the model becomes difficult if the information is redundant, irrelevant or unreliable. Exploring and cleaning the data becomes the most important and essential process in the development of a scorecard (Leung et al., 2008).

Fields or variables containing missing values tend to occur in industry, this can be attributed to a number of reasons, however there are techniques to handle them, and need to be used carefully. The best way to deal with missing values include the variables that have missing values in the scorecard. The missing value is seen as a distinct characteristic, that can be grouped and may be used as input; weights can then be assigned to this characteristic by the scorecard. Assigning a weight to the missing value may give insight into the behaviour of the missing values. Missing values are useful in the sense that it can be used as part of how the variable trends or it can be used as a bad indicator for performance. Therefore, it is recommended that the missing values be included in the analysis and be assigned scores in the scorecard (Siddiqi, 2006).

2.1.1 Initial characteristic analysis: Known Good Bad Model

To analyse the strength of the variable individually as a predictor, univariate screening is done. This is part of the data cleaning process which makes sure that a validation process is completed to check for the accuracy of the numerical codes for the values of each variable in the study i.e. remove all inefficient and irrational variables. Grouping is then applied to the strongest variables to obtain a certain format for the scorecard. Grouping variables has a few advantages such as, (1) outliers with interval variables and rare class become easier to work with, (2) easier to understand relationships, (3) grouping variables allows unparalleled control over the development process, (4) insight is gained into the behaviour of risk predictors and lastly (5) nonlinear dependencies may be formed with linear models (Siddiqi, 2006).

The strongest variables are grouped and selected, at the end of the process there will be a set of robust and grouped variables used in the modelling process.

Grouping is the most essential modelling step, where measures such as Weight of Evidence and Information Value are calculated.

The Information Value is a technique used in selecting significant variables in a predictive model. It ranks variables according to their importance. The Weight of Evidence is used to inform the predictive power of an independent variable with regard to the dependent variable (Bhalla, 2019).

2.2 Preliminary Scorecard: Known Good Bad model (KGB)

The initial characteristics analysis is done to identify a set of powerful variables that should be considered in the final model and will transform the variables into grouped variables. At this stage to select a set of characteristics several predictive modelling methods together may be used which may show the most predictive power. The model that is created at this stage is developed on accounts that have been accepted or declined for the product, thus the population is “known”. Generally, the KGB model consists of eight to fifteen variables to ensure that the predictive power will be strong irrespective of whether one or two variables were to be adjusted. At this stage, this process should produce a scorecard that consists of the best possible combination of variables, taking the following into consideration:

- Correlation between characteristics
- Final statistical strength of the scorecard
- Interpretability of characteristics at the branch/adjudication
- Implementable
- Transparency of methodology for regulatory requirements.

2.3 Reject Inference

Until this point all modelling done has been on accounts that have a known performance, known as the Known Good-Bad sample. When only using the approved (Known Good-Bad) population, numerous differences occur in every statistical analysis because of high sampling bias error. Reject inference is done to analyse the performance of previously rejected applications to estimate their behaviour (Paragon Business Solutions, 2014). Scorecard monitoring is based on the population that is accepted, thus, reject inference is used to predict the rejected applicant’s payment behaviour if the applicant had been granted credit.

By not taking the rejects into account the scorecard that would be generated would not be valid for the entire applicant population, thus reject inference is done. If a scorecard is built by making use of only the known goods and bads, it will predict that those who have delinquencies are good credit risks and should be approved credit. Reject inference “neutralizes” the distortive effects of “cherry-picking”; this can also be done for policy rules.

Reject inference generated a more realistic and accurate expected performance outcome for all applicants, from a decision-making perspective. Estimating the bad rates by score of the previously rejected applications creates better prospects for future performance by identifying a “swap set”; the swap set is the data set where the known bads are swapped for the inferred

goods. Inferred goods are applications that have been previously rejected and are now classified as possible goods after reject inference, in the future these applications will be accepted, together with reducing the known bads, that is swapping the known bads for the inferred goods; this will allow for the approval of the same number of applications while achieving an improved performance through “better-quality” selection.

2.4 Final Scorecard: All Good Bad model (AGB)

The final scorecard known as the All Good Bad (AGB) model is developed by running the matching initial characteristic analysis and statistical algorithms on the post-inferred dataset, that is the dataset after reject inference is done; to produce the final variables for the scorecard. Once a final scorecard is developed, further issues need to be dealt with such as scaling the scores, accuracy of the points allocated, misclassification and strength of the scorecard.

Scaling is the layout and range of scorecard and the rate of change in odds as the score increases. Scaling has no impact on the predictive strength of the scorecard, it's a decision function based on how the scorecard is implemented into the application software, the staff's ability to understand, and the stability with existing scorecards in the organisation.

2.4.1 Assignment of Points

The points allocated to each variable and the strength of the scorecard must be evaluated after the final scorecard has been developed. When assigning scores, it should make sense and follow trends that were recognized in the initial characteristic analysis. When scores are being allocated to characteristics it is selected in such a way that it is “prediction neutral”, it will not be bias to any applicant. If a reversal in the points allocated occurs, an analytical adjustment of the points is usually done, given that the rest of the points allocation is logical and there are no reverse trends. Regrouping may be needed; this will be dependent on how severe the reversal and order of the points assigned for the rest of the variables are affected. This is repeated until the scorecard is statistically and operationally acceptable (Siddiqi, 2006).

2.5 Validation

Validating the scorecard is essential once the final model has been selected. Validation is done to verify that the model that was developed is valid to the subject population, and to verify that model has not been overfit. A selection of techniques and tests are done to evaluate if the ratings adequately differentiate risk and if the estimates of risk factors, namely, PD, LGD and

EAD characterize the relevant factors of risk appropriately. PD is the expected probability of default, LGD is the loss given default in other words it is the portion of the asset that is lost if a customer defaults , and EAD is the exposure at default in other words it is the total value the credit granter is exposed to when a customer defaults. It's suggested that modelling be done with a 70% or 80% development sample and the remaining 30% or 20% hold-out sample be set aside for validation (Siddiqi, 2006).

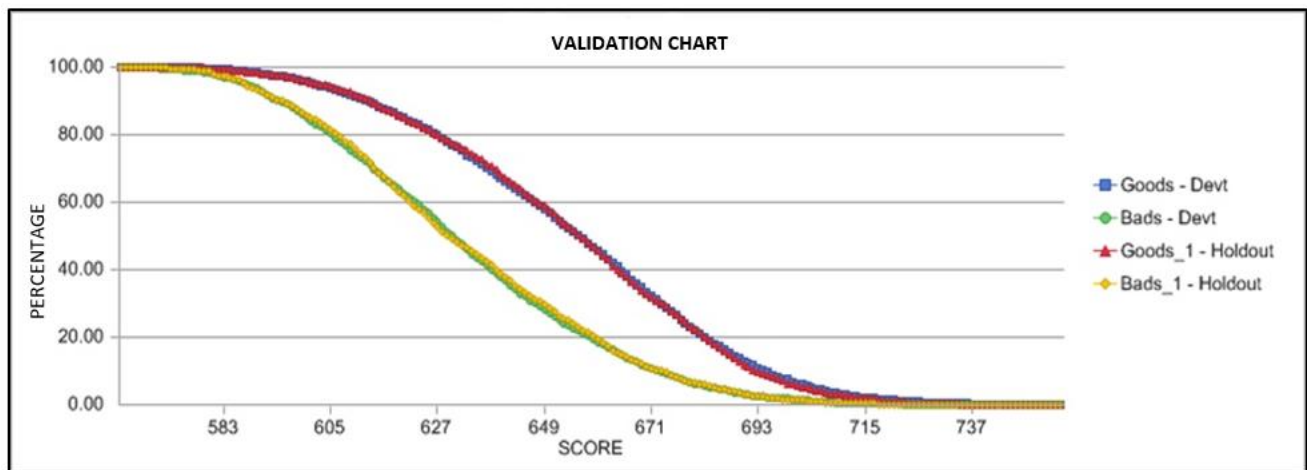


Figure 2. 2: Validation chart

Figure 2.2 illustrates an example of a validation chart. The scorecard is considered to be validated if there is no significant difference between the development sample and the holdout sample. Validation is done to confirm if the model has been overfitted or not. During model development it is good practice to model with a random 70% or 80% of the development sample, while the remaining 30% or 20% which is known as the holdout sample, is used for validation. The above chart shows the distribution of scored goods and scored bads across the holdout sample and the development sample. If the curves of the holdout sample and development sample have a great deviation from each other then the model may be overfitted.

2.5.1 Scorecard stability report

System stability report is often referred to as the “population stability index (PSI)” or “scorecard stability report”; the index measures the significance of the population shift between expected and recent applicants and is calculated as follows:

$$\sum (\%Actual - \%Expected) \times \ln(\%Actual/\%Expected) \quad (2.1)$$

In general:

- PSI < 0.10 shows no significant change
- 0.10 < PSI < 0.25 shows a small change that may need to be investigated
- PSI > 0.25 shows a significant shift in the applicant population

A shift in the population can occur due to an independent change in the applicant's profile, incorrect data capturing or errors in coding such as including exclusions in the development sample.

2.5.2 Variable analysis report

The characteristic analysis report gives more detail on the shifts in the distribution of the scorecard characteristics, and the influence on the scores because of that shift. The index is calculated as follows:

$$\sum (\%Actual - \%Expected) \times Points \quad (2.2)$$

Characteristic analysis is done to meticulously find the causes for the score shifts.

2.6 Post-Implementation

Once the scorecard is implemented on the system, scorecard and portfolio management reports that are created is as follows (Siddiqi, 2006):

Scorecard management reports are done to:

- to track incoming applications. The report that is run is the system/population/scorecard stability report.

Portfolio performance reports are done to:

- analyse the delinquency performance of the accounts. The reports that are produced are the delinquency report and delinquency migration report.

Scorecard and application management reports are done to:

- to confirm if “the future is like the past”. Scorecards are built for a particular applicant profile and needs to be validated continuously. The reports that are produced are the system “stability report, scorecard characteristic analysis and non-scorecard characteristic analysis”.
- observe and identify causes of change in the profiles of applicants. Recognizing that a change has occurred is inadequate. The cause of the change must also be identified. The reports that are produced are the “scorecard and non-scorecard characteristic analysis, analysis of competition and marketing campaigns and analysis by area and other segments”.
- track risk the profile of incoming customers and applicants. The reports that are produced are the “system stability report, scorecard and non-scorecard characteristic analysis and score distribution of approved customers report”.
- produce statistics for acceptance and overrides. The reports that are produced are the “final score report and override report”.

Portfolio management reports are done to:

- observe risk performance of accounts. The reports that are produced are the “delinquency report, vintage analysis, delinquency migration report and roll rate across time report”.
- observe and identify the causes of delinquency and revenue. Understanding where the shortfalls are coming from lets one to take risk-adjusted decisions. The reports that are produced are the “delinquency report, by region and other segments and marketing campaign and competitive analysis”.

- estimate potential loss rates. The report that is produced is the “vintage analysis and roll rate report”.
- assess bad rate estimates and manage expectations. Tracking actual performance against expected performance permits for adjustments for potential loss forecasts. The reports that are produced are the vintage.

2.7 Summary

By using all the steps mentioned in the preceding chapter, will help develop a scorecard that is efficient and meets the institutes requirements. The most important step in the process is the data exploration, the data needs to be cleaned and makes sense in order to for the scorecard to be as strong as possible if the data is not cleaned and does not make sense then the scorecard would not predict the required outcome i.e. “garbage in, garbage out”.

CHAPTER 3: DATA SOURCES AND PERFORMANCE DEFINITIONS

Chapter 3 will cover the data sourcing and analysis step. This chapter explains the different data sources used to extract the application and performance data. After extracting the required data, the performance definition was determined by defining the sample period and outcome period and doing a roll rates analysis. Once this is completed the modelling development process begins. The scorecard that will be developed will be for Revolving loans and Overdrafts combined.

3.1 Introduction

The chapter aims to provide high-level information surrounding the extraction of development data. The scorecard that will be built will be for Overdrafts and Revolving Loans facility for a financial institute. The model that will be built will be for existing customers applying for an Overdraft or Revolving Loan facility or for a limit increase on a current Overdraft or Revolving Loan facility. The scorecard will determine the risk at application stage of applicants applying for an Overdraft or Revolving Loan and/or limit increase by combining demographic information captured at time of application with internal company behavioural information and bureau data were available at time of application.

Data was extracted from relevant tables for application, affordability and performance data. The necessary joins were made using a unique key to join the relevant tables. Duplicates were removed from the final table/sample created.

The sample generally contains $\pm 10\%$ of Good, Bad, Indeterminate and Reject applications and should consist of at least 5000 records for each. In instances where 5000 records are not available then all available records should be included in the sample.

Sampling is then conducted in SAS using the SurveySelect procedure with the GBI definition (indicating whether the record is Good, Bad, Indeterminate, Reject or Not Taken Up) specified as the Strata and then using Simple Random Sampling for each definition in the Strata. Specifying the Strata statement within the SAS procedure partitions the dataset into non-overlapping groups defined by the GBI definition variable. Independent samples are then selected from these strata. Simple Random Sampling selects units with equal probability and without substitution, meaning that a unit cannot be chosen

3.2 Data sources

3.2.1 Application data source

Application data for all cheque account applications were obtained from the Application Database for the period 01 May 2013 to 31 October 2014.

Account specific data was also extracted from the performance tables for relevant accounts for each month in which the application took place.

Behavioural Risk Scoring data relevant at the time of the application was also extracted to obtain internal behavioural variables to be used in the modelling process.

3.2.2 Account performance data source

Revolving Loans and Overdrafts performance data was extracted from the performance tables for each month following the month in which the application took place up to September 2013.

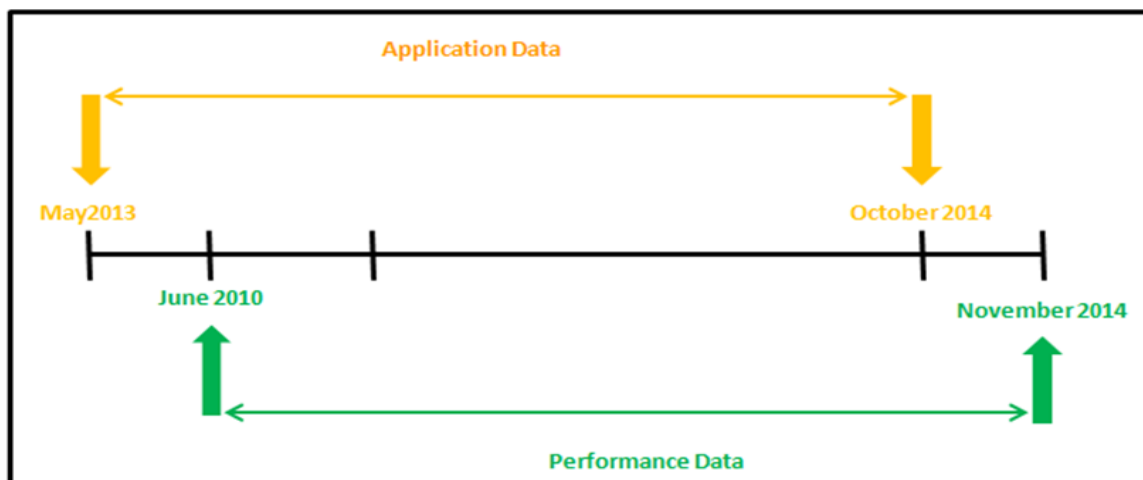


Figure 3. 1: Data extraction periods

Due to the fact that a customer may have multiple applications the most relevant application was kept according to the hierarchy:

- Approved – applications that have been approved for an overdraft/revolving loan
- Declined – applications that have been declined for an overdraft/revolving loan
- Pending – applications that are still pending an outcome
- Referred – these applications have not been declined or accepted
- Missing
- Other

The application data is the data used in the sample period, the performance data is the data used in the outcome period (Figure 1.3).

3.3 Performance definition

3.3.1 Good/bad/indeterminate/exclusion definition

The following definitions were utilised as based on the relevant days in arrears fields and was consistent for both Revolving Loans and Overdrafts.

Days in Excess	Good/Bad
0	Good
1-30	Good
31-35	Indeterminate
36-90	Indeterminate
91-120	Bad
121-150	Bad
151-180	Bad
181-210	Bad
211-230	Bad
231-270	Bad
271+	Bad

Table 3. 1 Good/Bad definition – Overdrafts

In Table 3.1 it can be seen for Overdrafts, that if the customer was 0-30 days in excess then the customer is considered good, if the customer was 31-90 days in excess then the customer is considered indeterminate and if the customer was 91-271+ days in excess then the customer is considered to be bad for an overdraft.

Total Payments Past Due	Good/Bad
0	Good
1	Good
2	Indeterminate
3	Bad
4	Bad
5	Bad
6	Bad
7	Bad
8	Bad
9	Bad

Table 3. 2: Good/Bad definition – Revolving Loans

Table 3.2 shows that for Revolving Loans, if the customer had 0-1 total payments past due then the customer is considered to be good, if the customer had 2 total payments past due then the customer is considered to be indeterminate and if the customer had 3-9 total payments past due then the customer is considered to be bad.

3.3.2 Roll rate analysis

To determine the optimal Good/Bad definition a roll rate analyses was performed.

The roll rate was performed as follows:

- All analyses were done on an application level, so the same customer may appear multiple times;

The worst performance for the first 12 months on book was compared to the worst performance in the next 12 months for all written accounts. Written accounts are all accounts that are on book, that is used to deposit and withdraw from the facility being used.

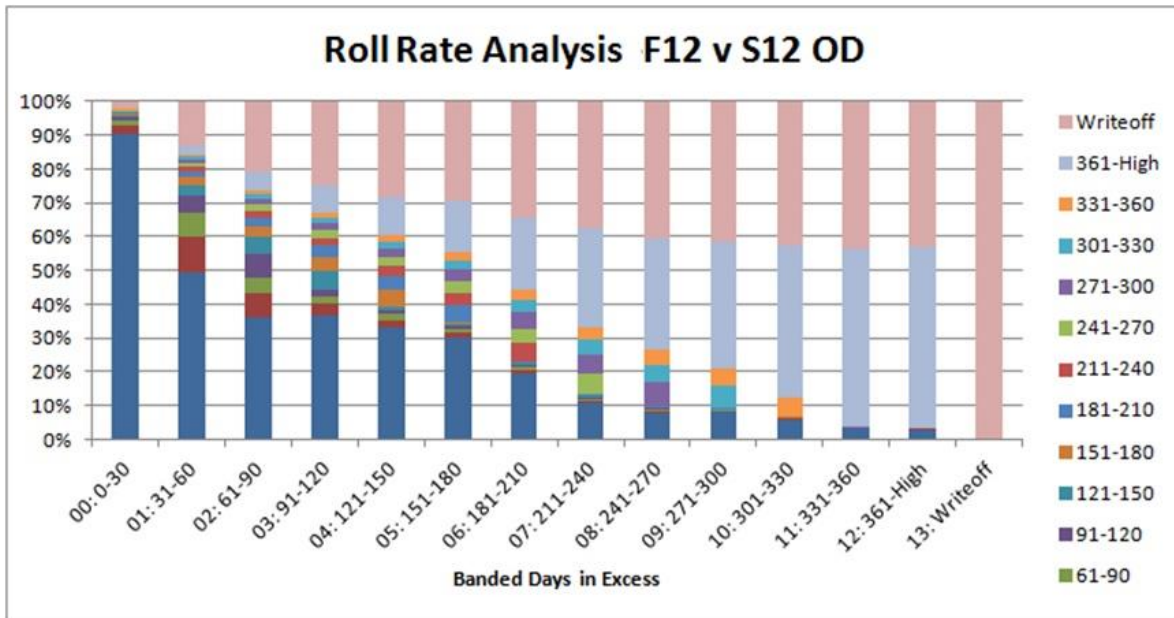


Figure 3. 2: Roll rates analysis 12 months - Overdrafts

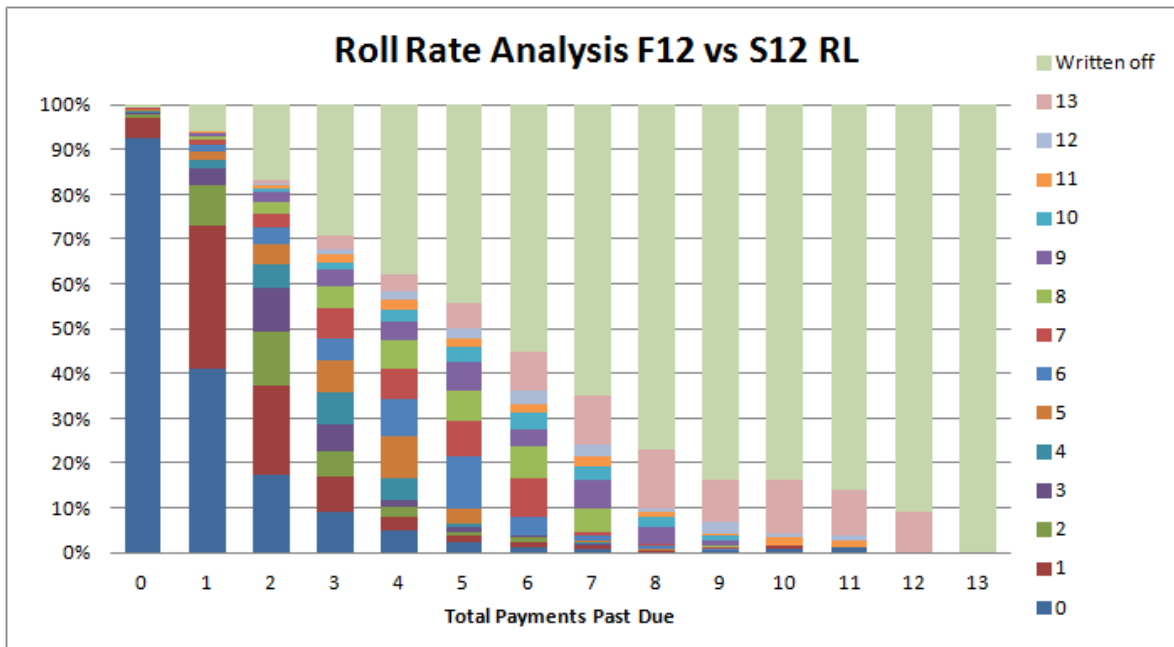


Figure 3. 3: Roll rates analysis 12 months – Revolving Loans

Roll rates analysis compares the worst delinquency status of account in previous number of 'n' months (e.g. previous 12 months) to the next number of 'n' months (e.g. next 12 months). The percentage of accounts that have stayed delinquent or gotten worse or better are then calculated. This method is used to determine if a customer will remain delinquent or move backward and forward between the delinquent statuses i.e. get better or worse.

From Figures 3.2 and 3.3, it can be seen that for Overdrafts, accounts that are 1 to 30 days in arrears in the first period, 87% remain in this status or improves in the second period. For Revolving Loans, accounts that are 91 to 120 days in arrears in the first period, only 40% roll to an improved status in the second period.

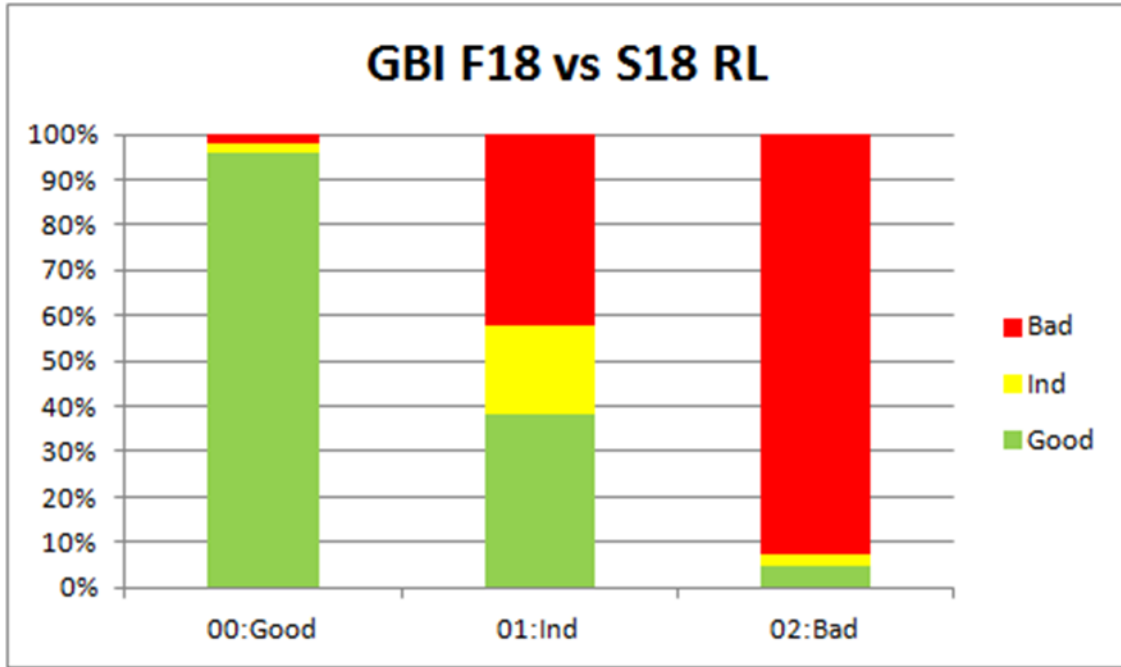


Figure 3. 4: Roll rate analysis 18-month definition – Revolving Loans

	%	Volume
% population that is Good:	94.43%	127038
% population that is Indeterminate:	2.51%	3373
% population that is Bad:	3.06%	4115
% Goods that remain Good:	96.02%	121977
% Goods that roll to Indeterminate:	1.96%	2488
% Goods that roll to Bad:	2.03%	2573
% Indeterminates that remain Indeterminate:	19.15%	646
% Indeterminates that roll to Good:	38.39%	1295
% Indeterminates that roll to bad	42.45%	1432
% Bads that remain Bad:	92.93%	3824
% Bads that roll to Indeterminate:	2.41%	99
% Bads that roll to Good:	4.67%	192

Table 3. 3: Roll rates analysis – Revolving Loans

From Table 3.3, for revolving loan accounts that were classified as Good in the first period, 90.02% remained in this state during the second period. Also 92.93% of accounts classified as Bad in the first period, remain Bad in the second period. The chosen Good/Bad definition is therefore stable and is considered adequate for scorecard development.

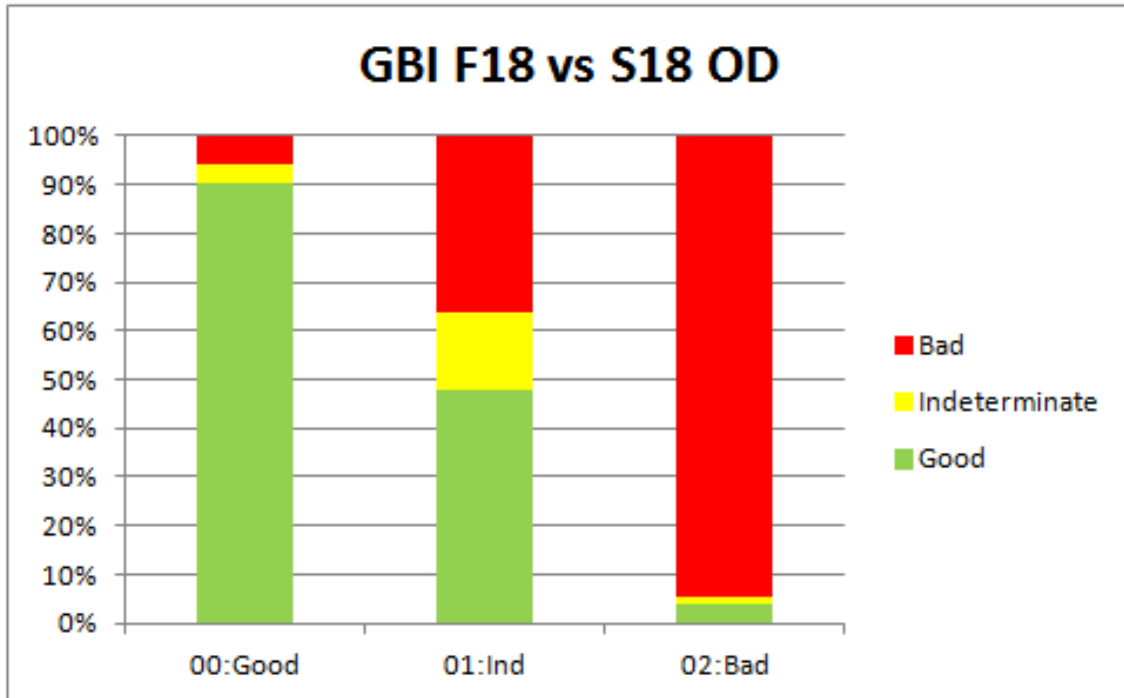


Figure 3. 5: Roll rates analysis 18-month definition - Overdrafts

In Figures 3.4 and 3.5 the roll rates analysis for Revolving Loans and Overdrafts, respectively, use an 18-month maturity period to obtain the Good/Bad definition. For Revolving Loans and Overdrafts, the Good/Bad definition is defined as, if the account is 0 payments in arrears then the client is deemed good, if the account is 1 payment in arrears then the client is deemed indeterminate and if the account is 2+ payments in arrears then the client is deemed bad. Thus, the overall definition used when developing the scorecard for Revolving Loans and Overdrafts is when the account is 0 payments in arrears then the client is good, when the account is 1 payment in arrears then the client is indeterminate and when the account is 2+ payments in arrears then the client is bad.

	%	Volumes
% population that is Good:	73.07%	1225654
% population that is Indeterminate:	6.85%	114897
% population that is Bad:	20.08%	336804
% Goods that remain Good:	90.34%	1107229
% Goods that roll to Indeterminate:	3.78%	46325
% Goods that roll to Bad:	5.88%	72100
% Indeterminates that remain Indeterminate:	15.88%	18248
% Indeterminates that roll to Good:	47.69%	54789
% Indeterminates that roll to bad	36.43%	41860
% Bads that remain Bad:	94.68%	318885
% Bads that roll to Indeterminate:	1.25%	4194
% Bads that roll to Good:	4.08%	13725

Table 3. 4: Roll rates analysis - Overdrafts

From Table 3.4, for Overdraft accounts that were classified as Good in the first period, 90.34% remained in this state during the second period. Also 94.68% of accounts classified as Bad in the first period, remain Bad in the second period. The chosen Good/Bad definition is therefore stable, consistent with Overdrafts and is considered adequate for scorecard development.

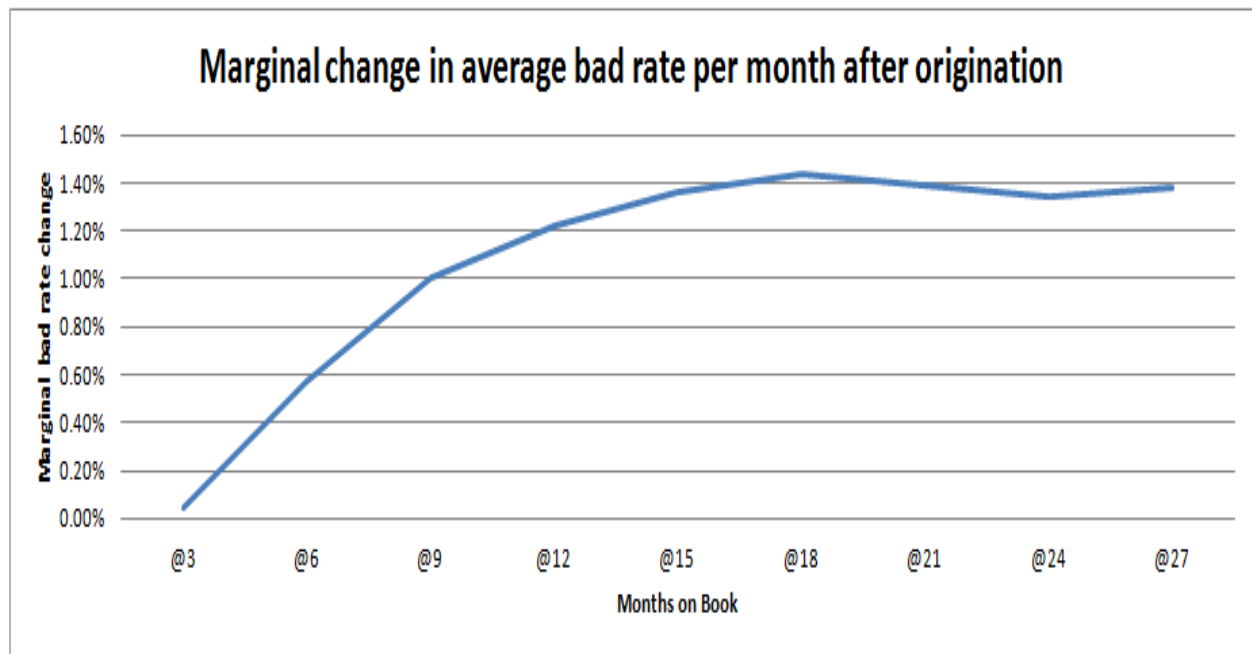


Figure 3. 6: Marginal bad rate change – Overdrafts

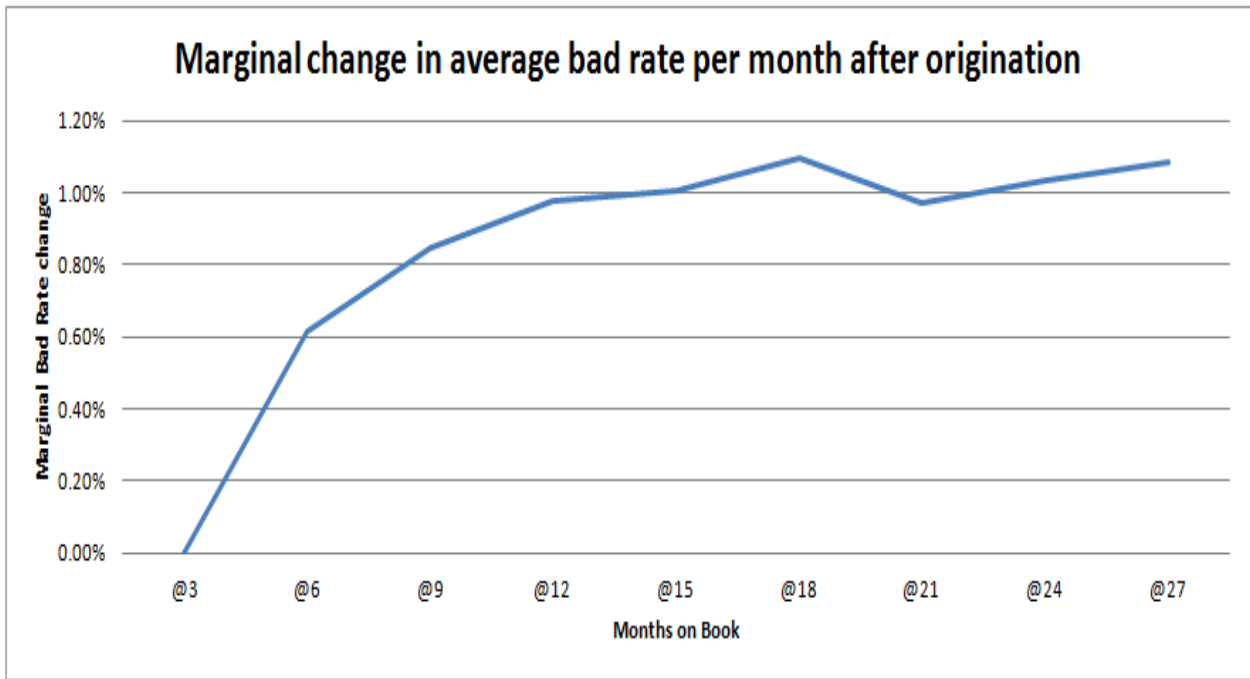


Figure 3. 7: Marginal bad rate change – Revolving Loans

Figures 3.6 and 3.7 shows the marginal bad rate peak suggests a peak occurring for both Overdrafts and Revolving Loans at 18 months. Though there was some elevation at higher periods. During the Masterfile analysis the viability of other outcome periods was tested. With regard to Masterfile overall, the 18-month definition wasn't explicitly clear, but its selection was done through balancing the investigation with the need for a reasonable period that product houses sufficed as adequate.

Development Sample

The development sample breakdown provided below. Note that the sample counts are indicated in the brackets.

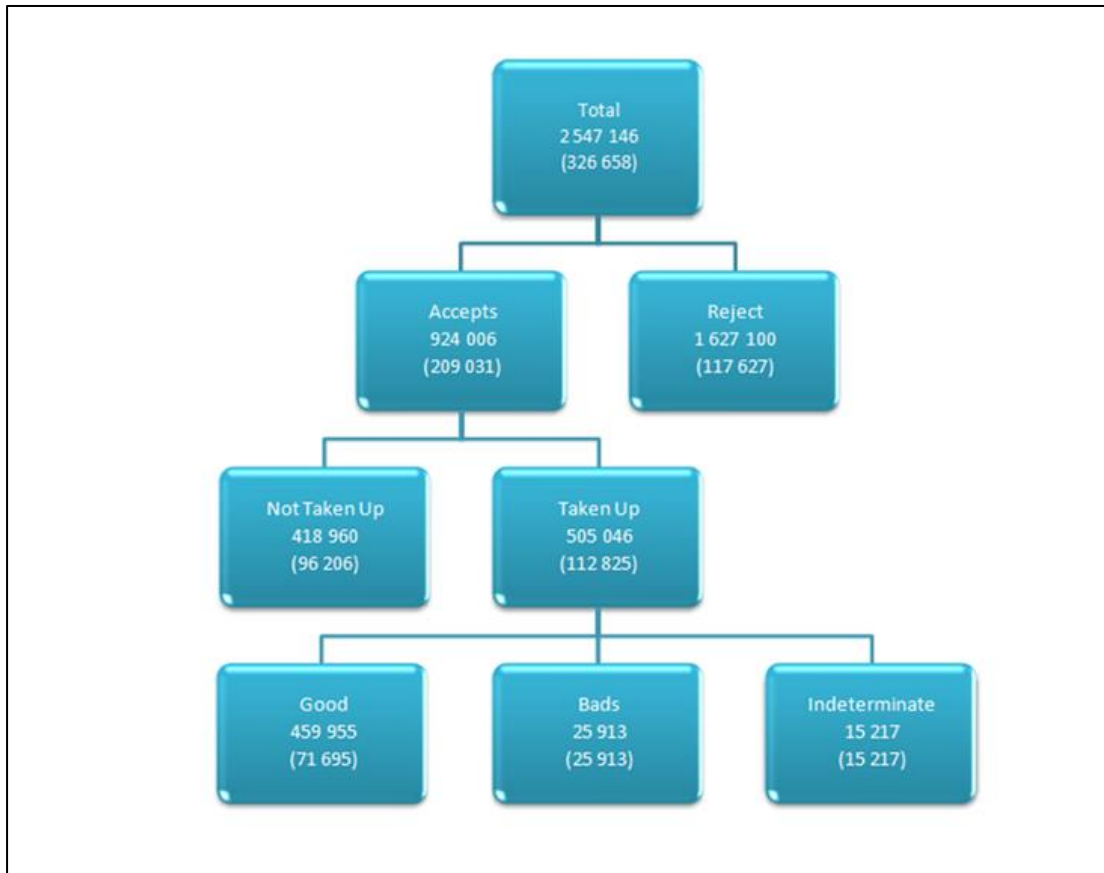


Figure 3. 8: Population flow - Development sample

Segmentation resulted in the development of 5 scorecards for the Overdraft and Revolving Loan portfolio. This study will focus on one segment, namely; existing to bank thick clients. Existing to bank customers are customers who are banking with the financial institute for more than six months and thick refers to the amount of bureau information that the institute has on the customer.

Sample Period

The sample period chosen was from 1 May 2013 to 31 October 2014. This was done based on the fact that portfolio takes 18 months to mature, maturity referring the time it takes for the outstanding loans to be repaid and also keeping in mind that we would have to perform an Out Of Time (OoT) validation can be done later on.

Figure 3.9 below shows the following:

- The sample window or observation period of 1 May 2013 to 31 October 2014;
- The outcome period, 1 November 2014 to 30 April 2016;
- Maturity of 18 months.

Note that the scorecard development was done on 80% of the sample with 20% used for the holdout validation.

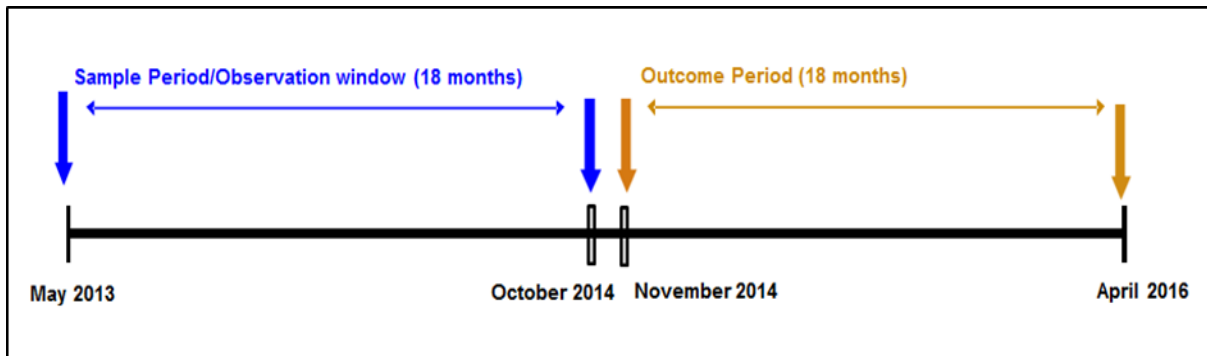


Figure 3. 9: Observation and Outcome period

A sample/observation period is the time period from which we sample when building a scorecard. Information on applicants is collected during a certain time period in the past. This information is used to build the scorecard. The outcome/performance period is the time period directly after the sample period; this is the time period allowed to observe the behaviour of the applicants that were accepted for credit during the sample period.

Population flow: Existing to Bank, Thick

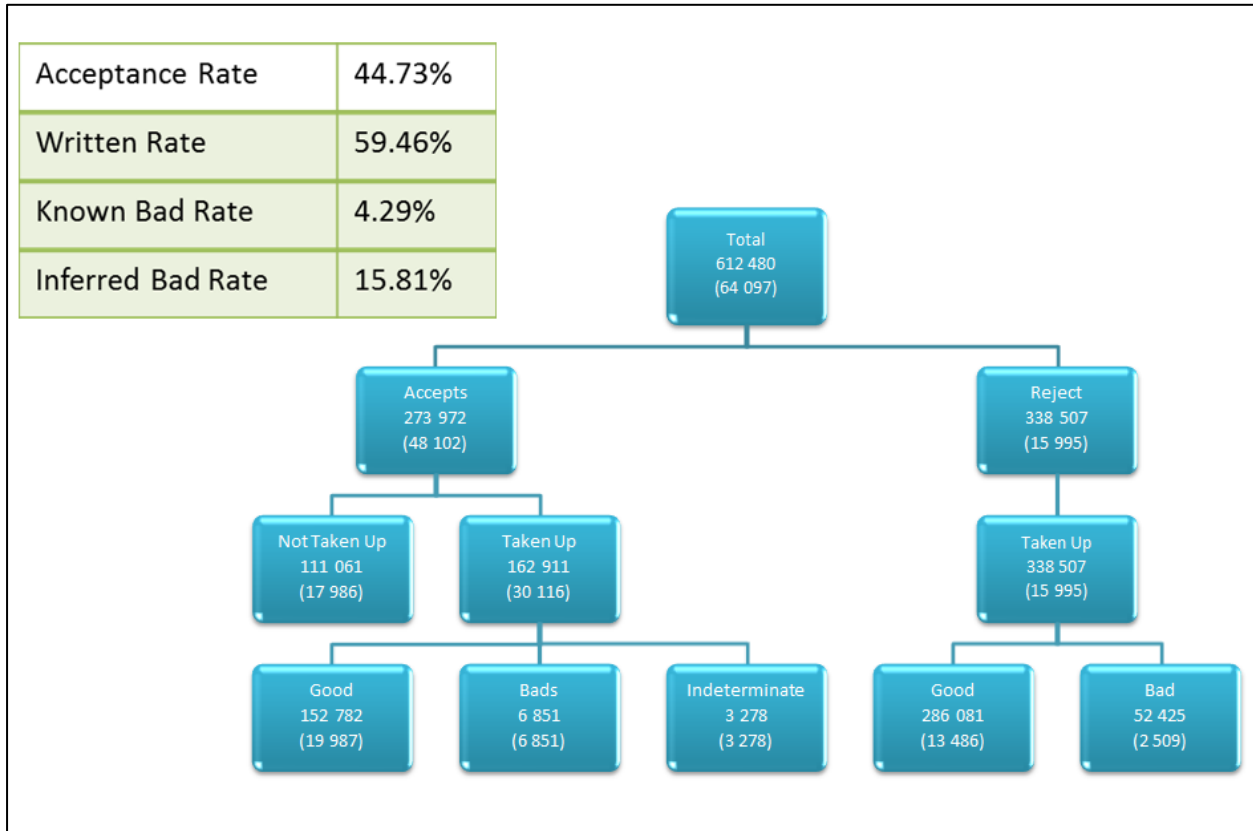


Figure 3. 10: Population flow Development sample - Existing to Bank, Thick

From Figure 3.10 it can be seen that a total population of 612 480 with 273 972 accepted customers and 338 507 rejected customers, the acceptance rate was 44.73%. From the accepted customers 111 061 customers had not taken up the product and 162 911 customers had taken up the product. From the accepted customers that had taken up the product; 152 782 were good customers, 6851 were bad customers and 3278 customers were indeterminate customers. From the rejected customers 338 507 had taken up the product and 286 081 of these customers were good and 52 425 were bad customers.

3.4 Summary

The spread between good, bad and indeterminate customers are adequate and enough to build a scorecard with a maturity period of 18 months. Thus, since the scorecard is developed for Overdrafts and Revolving Loans combined; when an applicant goes through the scorecard they will be applying for an Overdraft and a Revolving Loan, if the applicant is declined then the decline will be applicable for both products not one or the other.

Also, when developing a scorecard, one should consider the sample being used to develop the model on, the time windows being selected, the data sources being used as well as the quality of the data. Missing and marginal data need to be dealt with in an appropriate manner such that the results are not skewed.

CHAPTER 4: LOGISTIC REGRESSION

The following chapter will discuss the logistic regression and the assumptions used by it; the chapter will also discuss developing a scorecard using a logistic regression. The Gini coefficient will help when determining the strength of the predictive power of the models. The Known Good Bad model was first developed and validated. The Accepted Reject model was developed second and validated. Once the Known Good Bad model and Accept Reject model is developed, reject inference is then done. Once the reject inference is completed the final model known as the All Good Bad model is developed and validated; when looking at All Good Bad model the Weight of Evidence trends and stability over time trends are analysed.

4.1 Introduction

The logistic regression explores the “non-linear” association between a binary result and nominal or interval predictor variables. The outcome from logistic regression is the probability of an event occurring between 0 and 1. Logistic regression models that have several predictor variables are frequently identified in a stepwise or stratified manner, where the influence of each predictor can be evaluated by explaining the variation in the result. Not all predictors enhance the model fit and do not need to be in the model. Once the best possible model is found, it yields a mean log-odds for the likelihood to achieve one of the two possible outcomes which is dependent on the predictors in the model.

4.2 Assumptions of a logistic regression

The dependent variable is dichotomous in nature, that means that the event is either occurred or did not occur. There are no outliers present, this is achieved by converting the continuous predictors to uniform scores. There should be no correlation present between the predictors, a correlation matrix is constructed between the predictors, the rule generally used when building scorecards, no correlation of more than 0.5 is allowed. The logistic regression depends on analytically estimated coefficients, Euler’s constant and the log-odds transformation to predict probability (Koon and Petscher, 2015). The logistic regression allows for the estimation of multiple explanatory variables by expansion of the rudimentary principles. The general equation is denoted as (Eckel, 2008):

$$\text{logit}(p) = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \quad (4.1)$$

where:

p = probability event occurs.

The log-odds transformation is denoted as:

$$\text{odds} = \ln\left(\frac{p}{1-p}\right) \quad (4.2)$$

where:

p = the probability of an event happening

$1 - p$ = the probability of an event not happening

The log odds are the most applicable to use to model rather than probabilities because the log odds do not have the fundamental characteristic of the floor being 0 and the ceiling being 1 which is essential in probabilities; the probability that that an event will happen is between 0 and 1 and cannot be greater than 1 or less than 0. The log odds expand the scale to remove the effects of floor and ceiling boundaries. This is achieved by first taking the probabilities and transforming them to odds, and then take the log of the odds.

4.3 Logistic regression results

When using the logistic regression method, stepwise selection is used with the points to double odds set to 20, marginal odds set to 15 and odds to score set to 660. Scaling does not influence the predictive strength of the scorecard (Siddiqi, 2006).

Logistic regression models with multiple predictor variables are specified in a stepwise regression, where the evaluation of the influence of each variable is explained in the differences in the result. Predictor variables that do not statistically improve the model are not kept in the model; when the best fitted model is reached the log-odds of the likelihood is calculated by the model.

The stepwise technique which is mostly used for characteristic selection in regression models will always add the most influential obtainable characteristic at any step. In most conditions this

works fine. However, there are situations when this does not yield the most appropriate model, e.g.:

- A specific characteristic, or a small group of characteristics are significantly more powerful than the rest and influence the model such that added characteristics do not significantly better the model. This may occur when an external bureau score was used as a scored characteristic. Or,
- The presented characteristics are extremely correlated. Or,
- A combination of the above.

Multi-stage Modelling solution

- Dominant characteristics are kept back until the end, letting other characteristics to maximise their influence.
- Highly correlated characteristics are added to an existing scorecard rather than floating in a free mix with the rest of the characteristics.

4.4 The Gini coefficient

The Gini coefficient or Gini is a statistic to measure the ranking ability of the scorecard as well as the variables in the scorecard, in other others, the predictive power of the scorecard and variables in the scorecard.

If the Gini is 0% then the scorecard cannot distinguish between good and bad applicants. If the Gini is 100% then the Gini can distinguish accurately between good and bad applicants. An average credit scorecard has a Gini in the range of 40%-60%, the average behaviour scorecard has a Gini in the range of 70%-80%; powerful variables may have a Gini of 25% (Lucas, 2004).

4.5 Known good bad (KGB) model

The Known Good Bad model is developed on accounts that have been accepted or declined for the product, thus the population is “known”. A KGB model is built with a population distribution that is adapted by using the previously classified accept/reject weights. The KGB allocates classes that are driven by $p(bad)$ by combining the rejects with the accepts.

4.5.1 KGB variables with Gini

Fourteen variables were selected to be in the final Known Good Bad model. The Gini for the Known Good Bad model is 50.2.

Gini	Variable	Description
9,51	var 1	Debt Trajectory Var 1
27,32	var 2	Debt Trajectory Var 2
12,15	var 3	Debt Trajectory Var 3
13,81	var 4	Payment Profile Var 1
16,92	var 5	Payment Profile Var 2
4,16	var 6	Payment Profile Var 3
24,79	var 7	Payment Profile Var 4
16,29	var 8	Payment Profile Var 5
13,42	var 9	Behavioural Var 1
12,37	var 10	Transactional Var 1
11,91	var 11	Behavioural Var 2
9,32	var 12	Behavioural Var 3
7,96	var 13	Transactional Var 2
10,27	var 14	Transactional Var 3

Table 4. 1: List of Known Good Bad model variables and Gini's

The list of variables with their Gini's shows the predictive power of each variable that was selected to be in the Known Good Bad model.

4.5.2 KGB correlation matrix

Correlation Matrix														
	var 1	var 2	var 3	var 4	var 5	var 6	var 7	var 8	var 9	var 10	var 11	var 12	var 13	var 14
var 1	1	0.07	0.01	0.24	0.25	-0.07	0.19	0	0.03	-0.02	0.02	0.07	0	0
var 2	0.07	1	0.5	0.09	0.1	0.22	0.03	0.14	0.04	0.06	0.05	0.06	0	0.03
var 3	0.01	0.5	1	0.05	0.05	0.19	0.01	0.04	0.02	0.04	0.05	0.03	0.01	0.01
var 4	0.24	0.09	0.05	1	0.38	-0.1	0.19	-0.02	0.04	0.01	0.05	0.06	0	0.03
var 5	0.25	0.1	0.05	0.38	1	-0.1	0.43	-0.19	0.07	0	0.14	0.19	-0.02	0.01
var 6	-0.07	0.22	0.19	-0.1	-0.1	1	-0.08	0.14	0	0.04	-0.01	-0.01	0.01	0.01
var 7	0.19	0.03	0.01	0.19	0.43	-0.08	1	0.17	0.08	0	0.1	0.1	-0.01	0.02
var 8	0	0.14	0.04	-0.02	-0.19	0.14	0.17	1	-0.03	0.05	-0.13	-0.08	0.04	0.04
var 9	0.03	0.04	0.02	0.04	0.07	0	0.08	-0.03	1	0.02	0.17	0.06	-0.01	0.02
var 10	-0.02	0.06	0.04	0.01	0	0.04	0	0.05	0.02	1	-0.09	0.01	0.46	0.15
var 11	0.02	0.05	0.05	0.05	0.14	-0.01	0.1	-0.13	0.17	-0.09	1	0.11	-0.15	-0.09
var 12	0.07	0.06	0.03	0.06	0.19	-0.01	0.1	-0.08	0.06	0.01	0.11	1	-0.02	0.01
var 13	0	0	0.01	0	-0.02	0.01	-0.01	0.04	-0.01	0.46	-0.15	-0.02	1	0.17
var 14	0	0.03	0.01	0.03	0.01	0.01	0.02	0.04	0.02	0.15	-0.09	0.01	0.17	1

Table 4. 2: Known Good Bad model correlation matrix

A correlation matrix is a table showing correlation coefficients between sets of variables. Each random variable (x-axis) in the table is correlated with each of the other values in the table (y-axis). This allows one to see which pairs have the highest correlation. The above correlation matrix is stable; there is no correlation of more than 0.55. The rule of 0.55 is used because a correlation of 0.55 indicates a strong relationship between variables which tells us that both variables are essentially predicting in the same manner. In the case where a correlation of 0.55 is present, the variable with a lower Gini and less buckets is removed.

4.5.3 KGB validation graph

The validation graph is used to compare the distribution of goods and bads in the hold out sample (the sample on which we test) and the development sample (the sample on which we build the model).

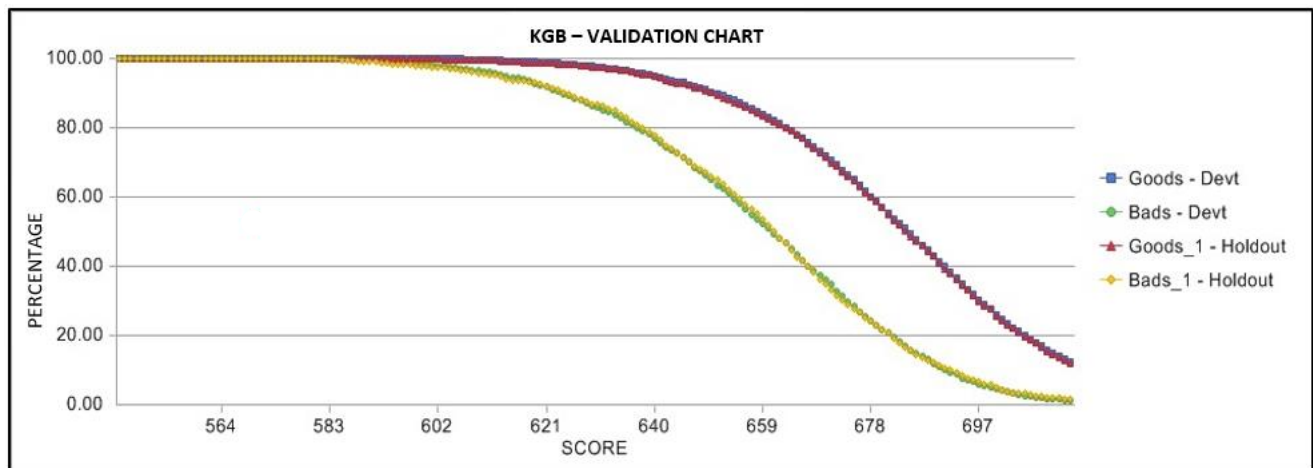


Figure 4. 1: Known Good Bad model validation graph

The holdout sample is a segment of the data that is not included in the modelling process and used in validation; the purpose of this is to make sure the scorecard predicts the outcome of the holdout sample accurately.

The validation chart in figure 16, trends exceptionally well, the two curves are not significantly different.

4.5.4 Variance inflation factors

The Variance Inflation Factor is a summary statistic indicative of multicollinearity within a model. A value of less than 2.5 for each characteristic is used as a maximum permissible value (Siddiq, 2006).

The Variance Inflation Factor for that variable is calculated as follows:

$$VIF = \frac{1}{(1 - R^2)}$$

(4.3)

The Variance Inflation Factor for the variables in each segment can be seen below:

VIF	Variable	Description
1,11	var 1	Debt Trajectory Var 1
1,43	var 2	Debt Trajectory Var 2
1,36	var 3	Debt Trajectory Var 3
1,22	var 4	Payment Profile Var 1
1,59	var 5	Payment Profile Var 2
1,1	var 6	Payment Profile Var 3
1,36	var 7	Payment Profile Var 4
1,2	var 8	Payment Profile Var 5
1,04	var 9	Behavioural Var 1
1,29	var 10	Transactional Var 1
1,11	var 11	Behavioural Var 2
1,05	var 12	Behavioural Var 3
1,31	var 13	Transactional Var 2
1,05	var 14	Transactional Var 3

Table 4. 3: List of Known Good Bad variables and VIF values

All Variance Inflation Factors are below the 2.5 threshold.

4.6 Accept reject (AR) model

The Accept Reject model is developed on the accepted and declined population following the same thought pattern as for the KGB. This model is predicting accept/reject and not good/bad.

4.6.1 AR variables with Gini

Three variables were selected to be in the final Accept Reject model. The Gini for the Accept Reject model is 65.1

Gini	Variable	Description
31,94	var 1	Affordability Var 1
14,45	var 2	Behavioural Var 4
63,35	var 3	Behavioural Var 5

Table 4. 4: List of Accept Reject model variables and Gini's

The list of variables with their Gini's, show the predict power of each variable that was selected to be in the Accept Reject model.

4.6.2 AR correlation matrix

Correlation Matrix			
	var 1	var 2	var 3
var 1	1	0.2	0.37
var 2	0.2	1	0.15
var 3	0.37	0.15	1

Table 4. 5: Accept Reject model correlation matrix

A correlation matrix is a table showing correlation coefficients between sets of variables. Each random variable (x-axis) in the table is correlated with each of the other values in the table (y-axis). This allows one to see which pairs have the highest correlation. The above correlation matrix is stable; there is no correlation of more than 0.55. The rule of 0.55 is used because a correlation of 0.55 indicates a strong relationship between variables which tells us that both variables are essentially predicting in the same manner. In the case where a correlation of 0.55 is present, the variable with a lower Gini and less buckets is removed.

4.6.3 Variance inflation factors

The Variance Inflation Factor is a summary statistic indicative of multicollinearity within a model. A value of less than 2.5 for each characteristic is used as a maximum permissible value (Siddiq, 2006).

The Variance Inflation Factor for that variable is calculated as follows:

$$VIF = \frac{1}{(1 - R^2)}$$

(4.4)

The Variance Inflation Factor for the variables in each segment can be seen below:

VIF	Variable	Description
1,18	var 1	Affordability Var 1
1,05	var 2	Behavioural Var 4
1.16	var 3	Behavioural Var 5

Table 4. 6: List of Accept Reject model variables and VIF values

All Variance Inflation Factors are below the 2.5 threshold

4.7 Reject inference

Reject inference is done to try and determine what would have happened to the rejects had they been accepted; in other words, would they have turned out to be good or bad re-payers? As they are rejects, it is likely that more of them would have become bad than actually happened with the accepts, however it is unlikely that all of them would have been bad (CSH, 2015).

Reject inference forms a very important part in scorecard development as it ensures that your new scorecard will be able to discriminate well between a good and a bad payer on the entire applicant base, and not only on previously accepted applications.

As a rule of thumb reject inference should always be performed except in the case where one of the below scenarios is applicable (Paragon Business Solutions, 2014):

- It has been proven that the previous decision process was extremely accurate, in other words we can prove that every single rejected customer would have turned out to be a bad re-payer;
- We can prove that the previous decision process was completely random;
- The previous acceptance rate was very high (90% and higher).

4.7.1 Reject inference process

The basic steps in the reject inference approach are detailed below:

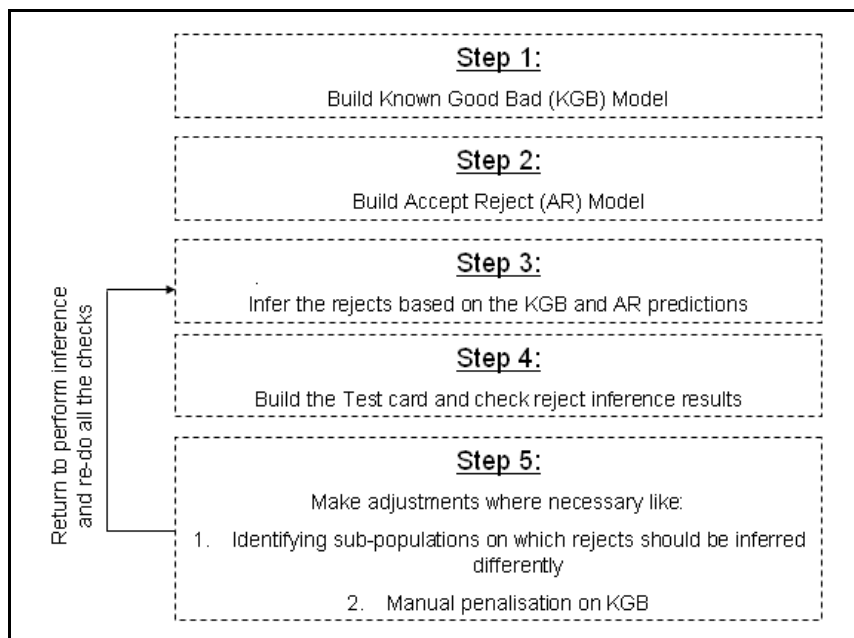


Figure 4. 2: Reject inference process

In a segment where the previous decision was strong, rejects are inferred closer to the higher benchmark (four times the known bad rate) to reinforce the previous decision. The converse is applicable for segments where the previous decision was weak.

The reject inference was conducted on the segment.

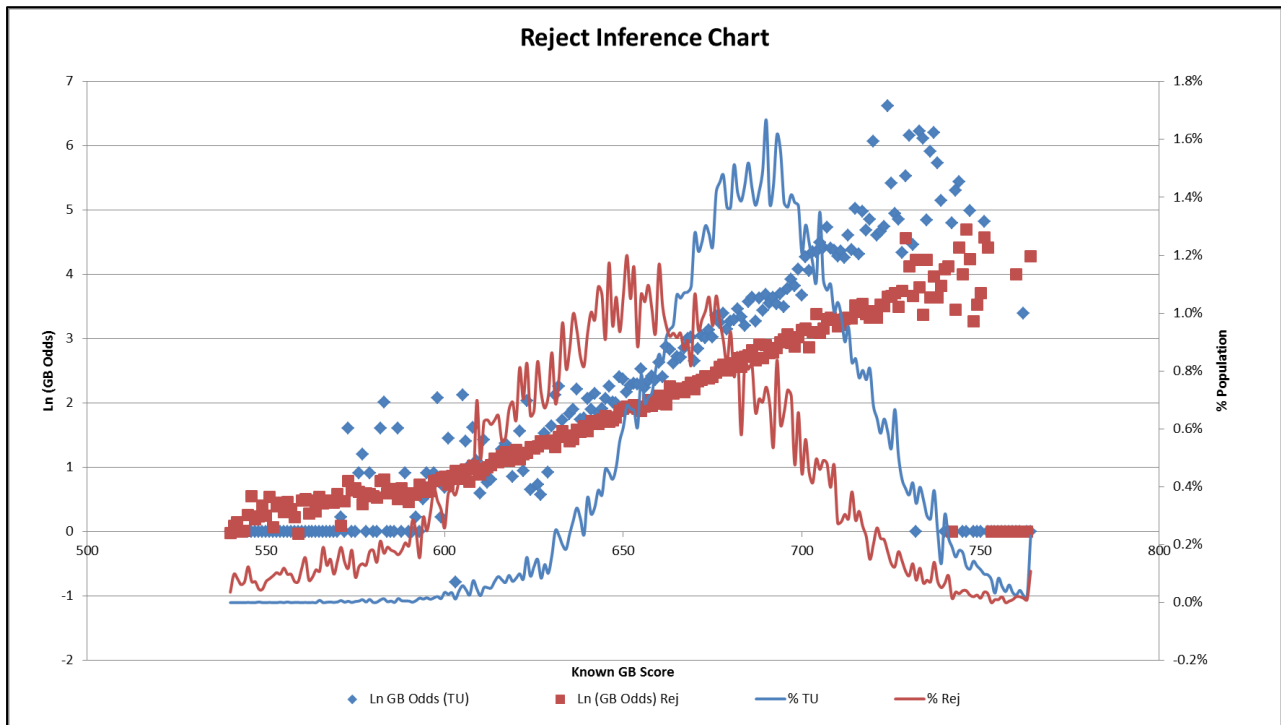


Figure 4. 3: Log odds chart - Reject inference results

The above chart known as log odds chart illustrates the occurrence in which rejects at a higher score range have a log odds line that diverges from the accept log odds proportionally more than rejects at a lower score range.

This occurrence implies that higher scoring rejects have been inferred more strictly than lower scoring rejects. This is quantified in the decreased ranking displayed by this segment in the post inference Gini value.

The above graph also indicates the following:

- The known goods and bads score higher than the rejects overall based on the new scorecard score;
- The log(odds) for the known goods and bads are better than the log(odds) for the rejects by the new scorecard score;
- The inference was done quite cautiously as the log(odds) for the rejects in the higher bands are quite a bit lower than for the known goods and bads.

While it is recognised that this is not optimal, the inference is still considered satisfactory for the following reasons:

- If rejects were to be swapped in, the swap in rejects would originate from the higher scoring rejects (which lie slightly below the cut-off at higher scores). As such, rejects that would in future be accepted have been inferred with a strict parameter which ensures that these applicants are correctly part of the accepted population;
- The occurrence can be partly ascribed to small volumes of rejects at that range (evidenced by the histogram distribution of records at that range) , reducing the ability of the modeller to control inference parameters at that range;
- The decrease in ranking is not considered material and the ranking of the model on the known population is preserved;
- Operational strategies are unlikely to increase cut-off points to a large enough extent for this anomaly to become material.

4.7.2 Reject inference checks

The first step after inferring the rejects is to build a Test card to measure the impact and the results of inference. The test card is based on assigning the same groups as for the KGB and including the same variables as were included into the KGB. This means we now have a like-for-like comparison.

Start by checking the bad rate for the known goods and bads (KGB), the Inferred Rejects and the overall adjustment made, assuming we had accepted all declined applications.

Bad rate improvements at same acceptance rate

The bad rate figures before and after reject inference are as follows:

	Thick ETB
Known Bad Rate	4.30%
Inferred Bad Rate	15.81%
Overall Bad Rate	11.70%

Table 4. 7: Bad rate adjustment

The rule of thumb is to confirm that the Inferred Bad rate is 1.5 to 4 times that of the known population. This is definitely the case for each segmented scorecard developed; however, overall the inferred bad rate is 4.4 times that of the known population. The bad rate figures are also provided for the child segments.

We then move on to segment specific checks and the next step is to measure the predicted improvement in bad rate at the previous acceptance rate:

	Thick ETB
Improvement after Reject Inference	33%

Table 4. 8: Bad rate improvement

The bad rate improvements for the reject inference segments usually require an improvement of between 5% and 30%. These do drift outside the margin, but the following was done to assess impact:

Post reject inference: we analysed the resulting Gini of “Dev excluding rejects” as a double check on the AGB model without rejects and confirmed that the RI process didn’t adversely influence our variable selection. We also noted that our rejects did score quite low and hence when inferred inflated our bad rates on the higher side of the 30% improvement guideline. The Gini for each reject inference segment is provided below:

	Thick ETB
Gini before Reject Inference	50.20%
AR Model Gini	65.10%
Gini after Reject Inference	56.60%
Change in Gini	12.75%

Table 4. 9: Gini improvement

Gini improvements did not exceed the 30% threshold for either reject inference segment.

4.8 All Good Bad (AGB) model

The All Good Bad Model (AGB) is the final model built. The sample used to build the AGB consists of the known goods, inferred goods and bads.

4.8.1 AGB variables with Gini

Eleven variables were selected to be in the final All Good Bad model. The Gini for the All Good Bad model is 56.

Gini	Variable	Description
11,52	var 1	Debt Trajectory Var 1
43,4	var 2	Debt Trajectory Var 2
22,16	var 3	Payment Profile Var 6
10,11	var 4	Payment Profile Var 3
29,34	var 5	Payment Profile Var 4
16,89	var 6	Behavioural Var 1
11,93	var 7	Behavioural Var 2
5,76	var 8	Behavioural Var 6
8,38	var 9	Behavioural Var 3
25,75	var 10	Transactional Var 1
18,39	var 11	Transactional Var 4

Table 4.11: List of All Good Bad Model Variables and Gini's

The list of variables with their Gini's shows the predictive power of each variable that was selected to be in the All Good Bad model. In the AGB model eight of the eleven variables were variables that were also in the final KGB model. The variables in particular are Debt Trajectory Var 1, Debt Trajectory Var 2, Payment Profile Var 3, Payment Profile Var 4, Behavioural Var 1, Behavioural Var 2, Behavioural Var 3 and Transactional Var 1.

4.8.2 AGB correlation matrix

<i>Correlation Matrix</i>											
	var 1	var 2	var 3	var 4	var 5	var 6	var 7	var 8	var 9	var 10	var 11
var 1	1	0.03	0.25	-0.08	0.23	0.03	-0.01	0.09	0.07	0.02	0.01
var 2	0.03	1	0.15	0.3	0.14	0.13	0.12	0.04	0.05	0.21	0.16
var 3	0.25	0.15	1	-0.23	0.39	0.12	0.13	0.01	0.1	0.08	0.05
var 4	-0.08	0.3	-0.23	1	-0.09	0.02	0	0.02	-0.03	0.09	0.07
var 5	0.23	0.14	0.39	-0.09	1	0.15	0.16	0.04	0.11	0.1	0.09
var 6	0.03	0.13	0.12	0.02	0.15	1	0.18	0	0.05	0.05	0.04
var 7	-0.01	0.12	0.13	0	0.16	0.18	1	-0.06	0.07	0.03	0.01
var 8	0.09	0.04	0.01	0.02	0.04	0	-0.06	1	0.05	0.02	0.02
var 9	0.07	0.05	0.1	-0.03	0.11	0.05	0.07	0.05	1	0.04	0.03
var 10	0.02	0.21	0.08	0.09	0.1	0.05	0.03	0.02	0.04	1	0.42
var 11	0.01	0.16	0.05	0.07	0.09	0.04	0.01	0.02	0.03	0.42	1

Table 4. 10: All Good Bad model correlation matrix

A correlation matrix is a table showing correlation coefficients between sets of variables. Each random variable (x-axis) in the table is correlated with each of the other values in the table (y-axis). This allows one to see which pairs have the highest correlation. The above correlation matrix is stable; there is no correlation of more than 0.55. The rule of 0.55 is used because a correlation of 0.55 indicates a strong relationship between variables which tells us that both variables are essentially predicting in the same manner. In the case where a correlation of 0.55 is present, the variable with a lower Gini and less buckets is removed.

4.8.3 AGB validation graph

The validation graph is used to compare the distribution of goods and bads in the hold out sample (the sample on which we test) and the development sample (the sample on which we build the model).

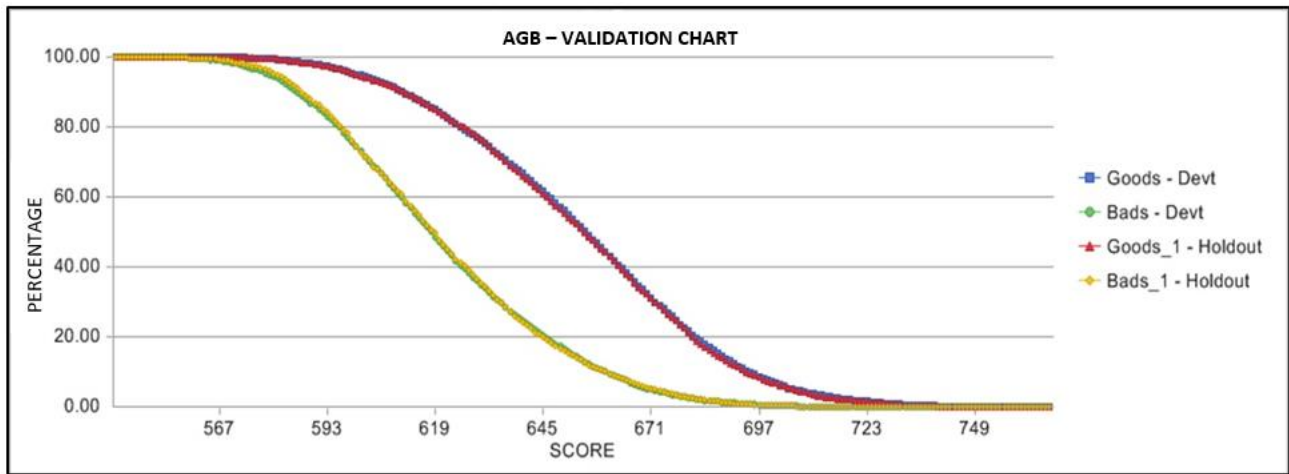


Figure 4. 4: All Good Bad model validation graph

In Figure 4.4, the holdout sample is a segment of the data that is not included in the modelling process and used in validation; the purpose of this is to make sure the scorecard predicts the outcome of the holdout sample accurately.

The validation chart in Figure 4.4, trends exceptionally well, the two curves are not significantly different.

4.8.4 Variance inflation factors

The Variance Inflation Factor is a summary statistic indicative of multicollinearity within a model. A value of less than 2.5 for each characteristic is used as a maximum permissible value (Siddiq, 2006).

The Variance Inflation Factor for that variable is calculated as follows:

$$VIF = \frac{1}{(1 - R^2)}$$

(4.5)

The Variance Inflation Factor for the variables in each segment can be seen below:

VIF	Variable	Description
1,11	var 1	Debt Trajectory Var 1
1,23	var 2	Debt Trajectory Var 2
1,33	var 3	Payment Profile Var 6
1,2	var 4	Payment Profile Var 3
1,25	var 5	Payment Profile Var 4
1,06	var 6	Behavioural Var 1
1,08	var 7	Behavioural Var 2
1,02	var 8	Behavioural Var 6
1,02	var 9	Behavioural Var 3
1,25	var 10	Transactional Var 1
1,22	var 11	Transactional Var 4

Table 4. 13: List of All Good Bad variables and VIF values

All Variance Inflation Factors are below the 2.5 threshold.

4.9 Weights of Evidence trend

In a Weights of Evidence model, the value of the variables used in the model build are not the characteristic values themselves, but the Weight of Evidence of the group to which it belongs, e.g. for accounts with age in the range 21-22, value -0.8648 would be used in the regression analysis (Paragon Business Solutions, n.d.).

The model returns results in the form:

$$0 = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_nx_n + E \quad (4.6)$$

Where

β_0 = intercept value

β_n = nth coefficient

x_n = nth Weight of Evidence value

E = residual Error

The model coefficients (Beta values) are multipliers to be applied to the WoE values. Logistic models are typically scaled to give a pre-determined Marginal Bad Rate (MBR) or Odds (D) at a given score (S) and a scaling factor, expressed as Points to Double the Odds (PDO). To convert from Score to Probability or vice-versa, we must first remove the scaling. In MODELLER, the logistic natural / raw score has been converted with the straight-line function (Paragon Business Solutions, n.d.):

$$Y = Mx + C \quad (4.7)$$

where:

Y = new points

$M = \frac{PDO}{\ln(2)}$

x = the raw score

C = scorecard constant

4.9.1 Weights of evidence graph

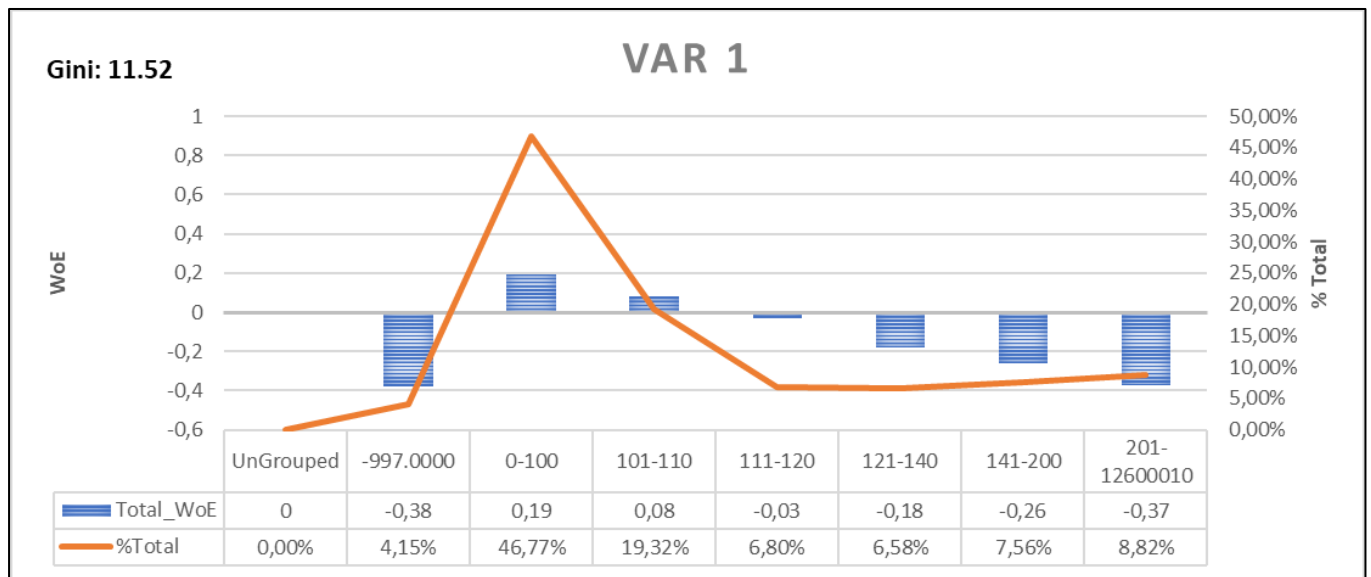


Figure 4. 5: WoE graph Var 1

The WoE graph is Figure 4.5 of Var 1 is first positive and then becomes negative indicating that the distribution of goods is higher between groups 0-120 than in groups >120. According to business logic the bad rate for this variable would increase as the ration increase, thus indicating the distribution of bads would be greater in subsequent groups. The ungrouped and group -997 which is known as a special code and is not taken into consideration in the trend due to no or low volumes.

In buckets -997 and “ungrouped” the negative WoE indicates predicting the outcome of applications in these buckets would be less likely, considering that bucket -997 is a special bucket, indicating missing information. Majority of the population fall in buckets 0-100 and 111-110 show the predictiveness in these groups is much higher and better at predicting the outcome of application if application falls within these groups.

The linearity across the WoE values shows that the credit default decreases across the different buckets. A monotonic relationship exists between the risk of defaulting and var 1 of an application. The WoE graph is used to show the subjective power of the input variable and to assess relationship between the input variable and the result.

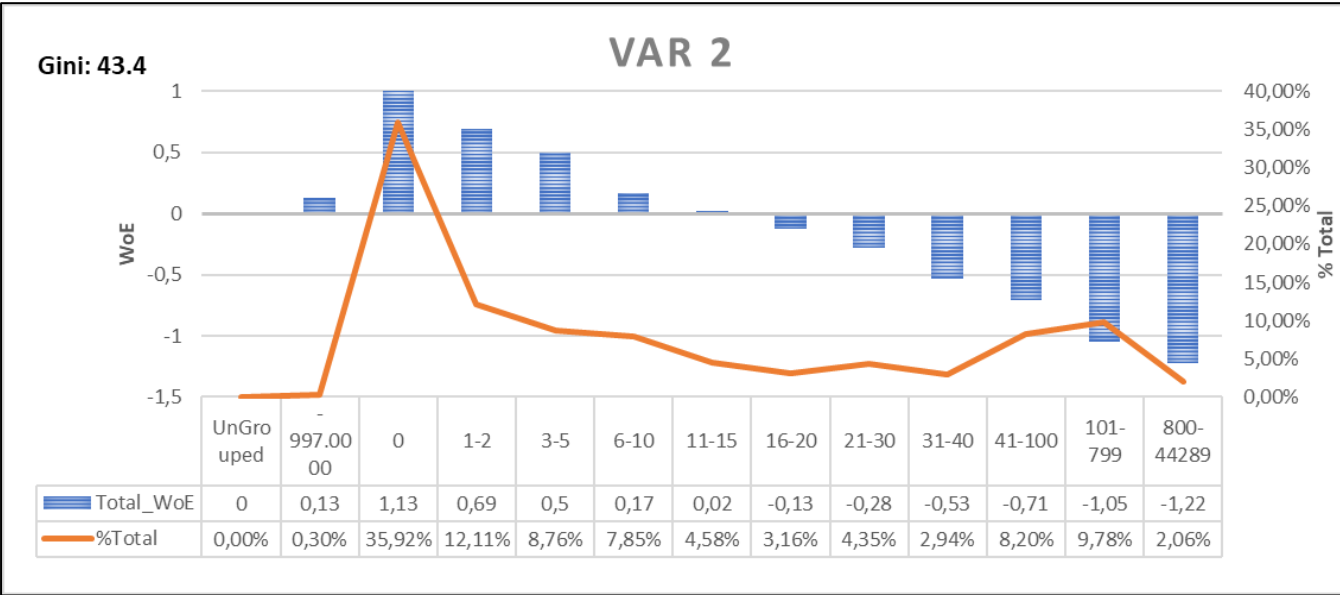


Figure 4. 6: WoE graph Var 2

The WoE graph in Figure 4.6 of Var 2 is first positive and then becomes negative indicating that the distribution of goods is higher between groups 0-15 than in groups > 15. According to business logic the bad rate for this variable would increase as the ration increase, thus indicating the distribution of bads would be greater in subsequent groups. The ungrouped and group -997 which is known as a special code and is not taken into consideration in the trend due to no or low volumes.

Var 2 displays a monotonic relationship between an application defaulting and Var 2, this shows that there is linearity present in the different buckets across Var 2. Greater part of the population falls in buckets 0-15 show high predictiveness in these groups and will be better at predicting the outcome of application if the application falls within these groups.

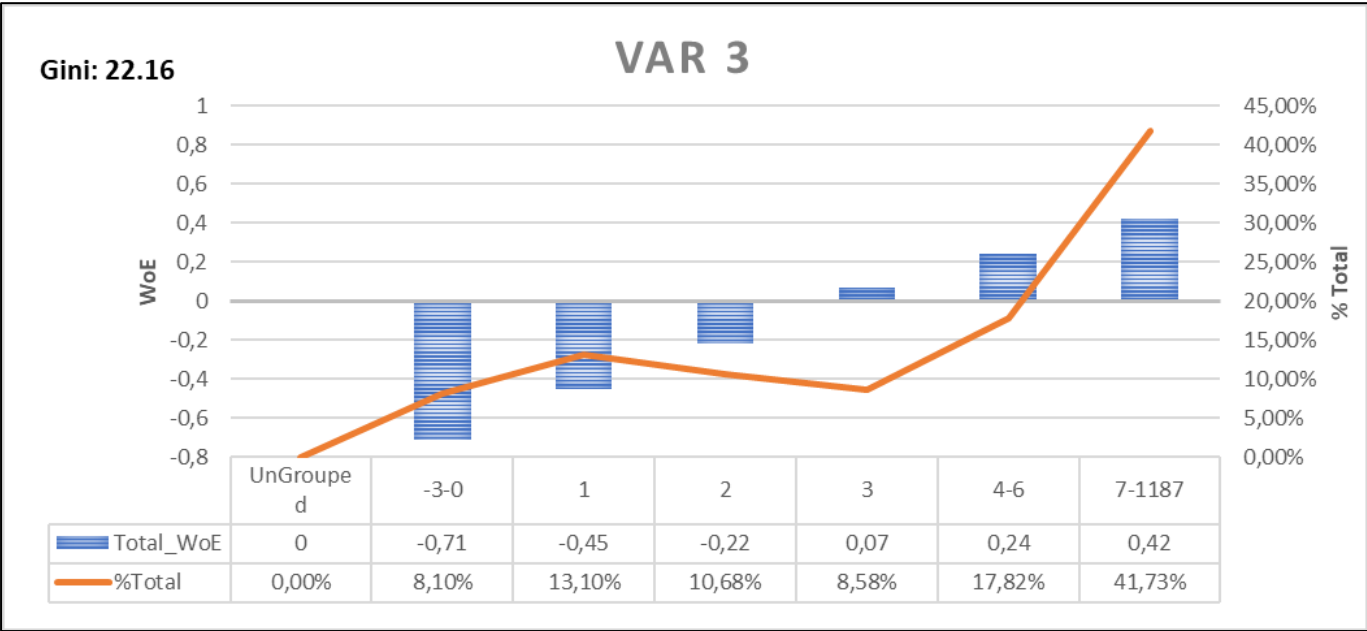


Figure 4. 7: WoE graph Var 3

The WoE graph in Figure 4.7 of Var 3 is first negative and then becomes positive indicating that the distribution of bads is higher between groups -3-2 than in groups > 2. According to business logic the bad rate for this variable would decrease as the months increase, thus indicating the distribution of goods would be greater in subsequent groups. The ungrouped is not taken into consideration in the trend due to no volumes.

Buckets -3-2 indicates the risk of default to be higher than in buckets 3-1187 as the WoE indicates that the that predicting the outcome in buckets 3-1187 would be higher as the predictive power would be stronger than in buckets -3-2.

The linear relationship between the buckets in the Var 3 indicate a monotonic relationship between the risk of default and Var 3.

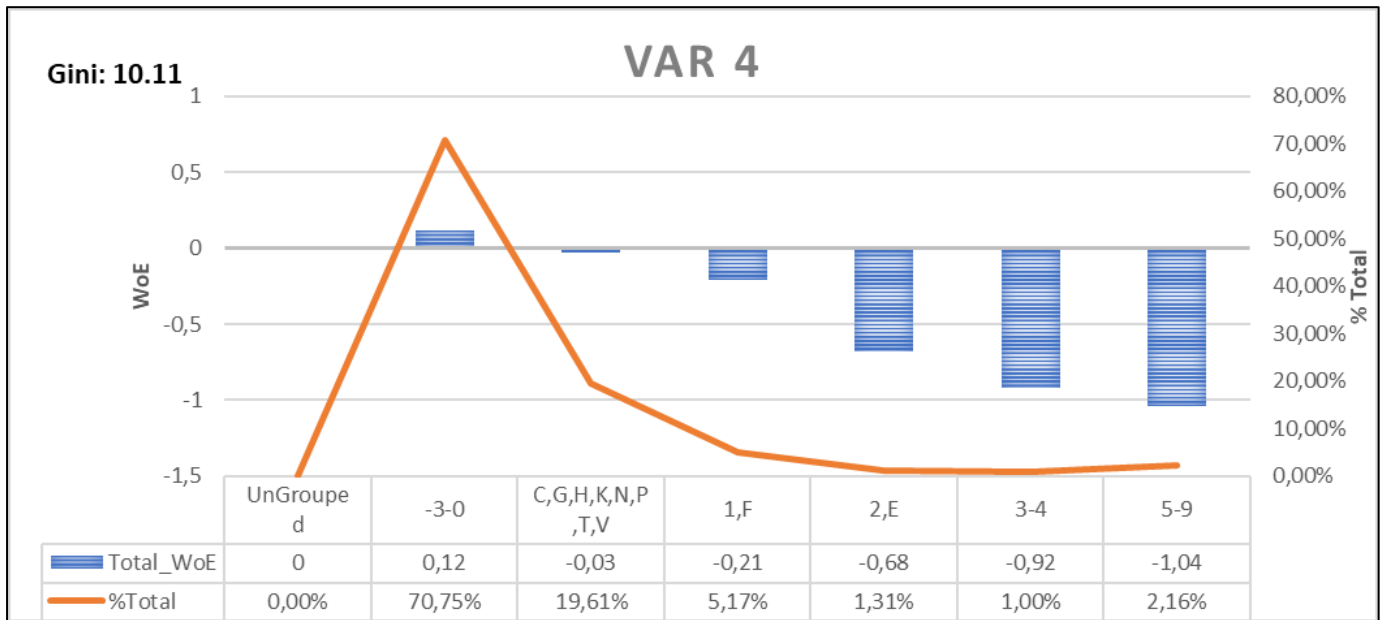


Figure 4. 8: WoE graph Var 4

The WoE graph in Figure 4.8 of Var 4 is first positive and then becomes negative indicating that the distribution of goods is higher between groups -3-0 and C,G,H,K,N,P,T,V than in groups 1,F,2,E,3,4 and 5-9. According to business logic the bad rate for this variable would increase as the status gets worse, thus indicating the distribution of bads would be greater in subsequent groups. The ungrouped is not taken into consideration in the trend due to no volumes.

The positive WoE for bucket -3-0 suggests applications would score better in this bucket than in the other buckets. The Negative WoE indicates an increase in defaults thus suggesting that applications that fall in these buckets would score worse as the prediction power would be lower in these buckets.

The monotonic relationship that exists between defaults and Var 4 indicates linearity between the WoE across the buckets in Var 4.

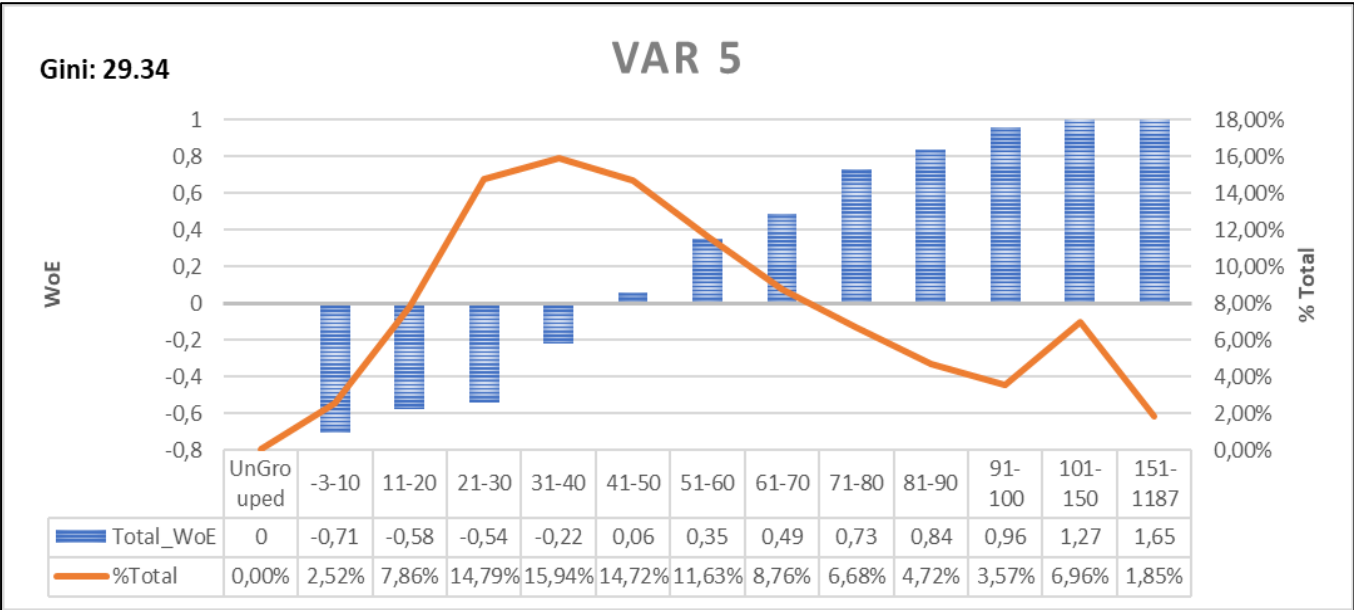


Figure 4. 9: WoE graph Var 5

The WoE in the above graph of Var 5 is first negative and then becomes positive indicating that the distribution of bads is higher between groups -3-50 than in groups > 50. According to business logic the bad rate for this variable would decrease as the age increase, thus indicating the distribution of goods would be greater in subsequent groups. The ungrouped is not taken into consideration in the trend due to no volumes.

Buckets 41-1187 have positive WoE indicating a strong predictive power in predicting the outcome as this indicates a decrease in credit defaults while buckets -3-40 indicate a weak predictive power in predicting the outcome as this indicates an increase in credit defaults.

The linearity across the WoE values shows that the credit default decreases across the different buckets. A monotonic relationship exists between the risk of defaulting and var 5 of an application. The WoE graph is used to assess relationship between the input variable and the result.

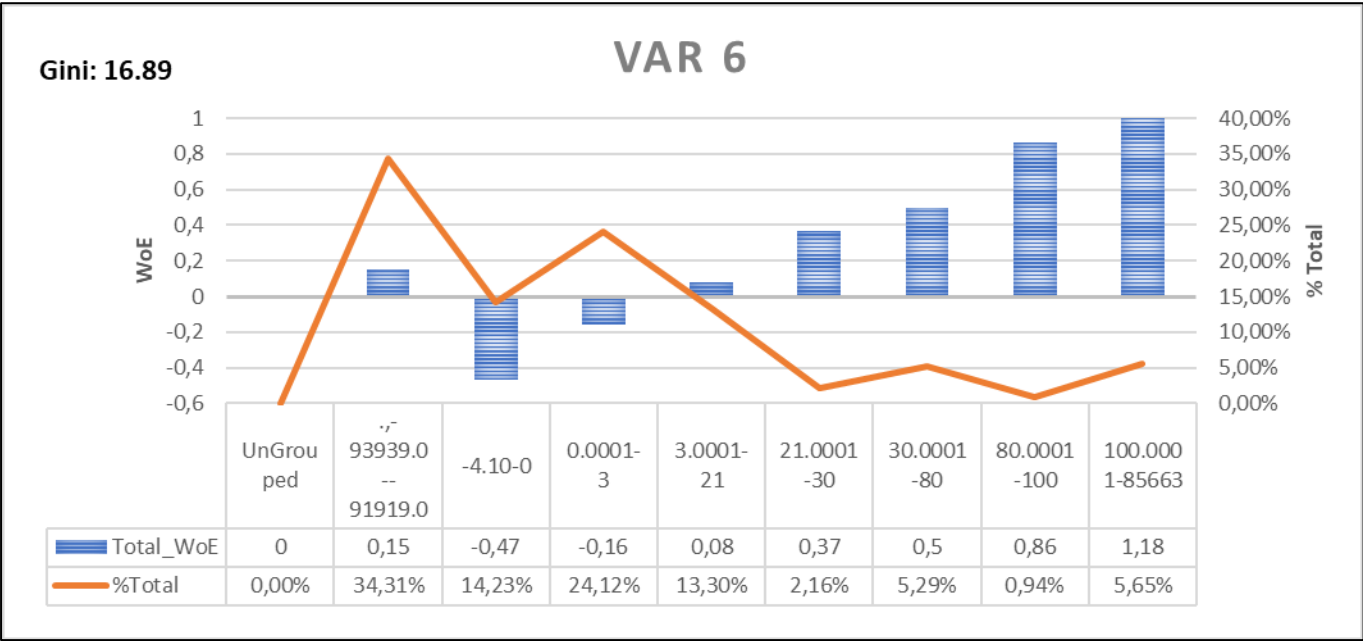


Figure 4. 10: WoE graph Var 6

The WoE graph in Figure 4.10 of Var 6 is first negative and then becomes positive indicating that the distribution of bads is higher between groups -3-3 than in groups > 3. According to business logic the bad rate for this variable would decrease as the ratio increases, thus indicating the distribution of goods would be greater in subsequent groups. The ungrouped is not taken into consideration in the trend due to no volumes and group., -93939.0--91919.0 is not taken into consideration in the trend because they are known a special code.

A monotonic relationship exists between the risk of defaulting and var 6 of an application. This indicates linearity across the WoE values which displays that the credit default decreases across the different buckets. The predictive power in buckets 3.0001-85663 are stronger than in buckets -4.10 – 3. The predictive power in buckets 3.0001 – 85663 indicates applications falling into these buckets are more likely to get a better score.

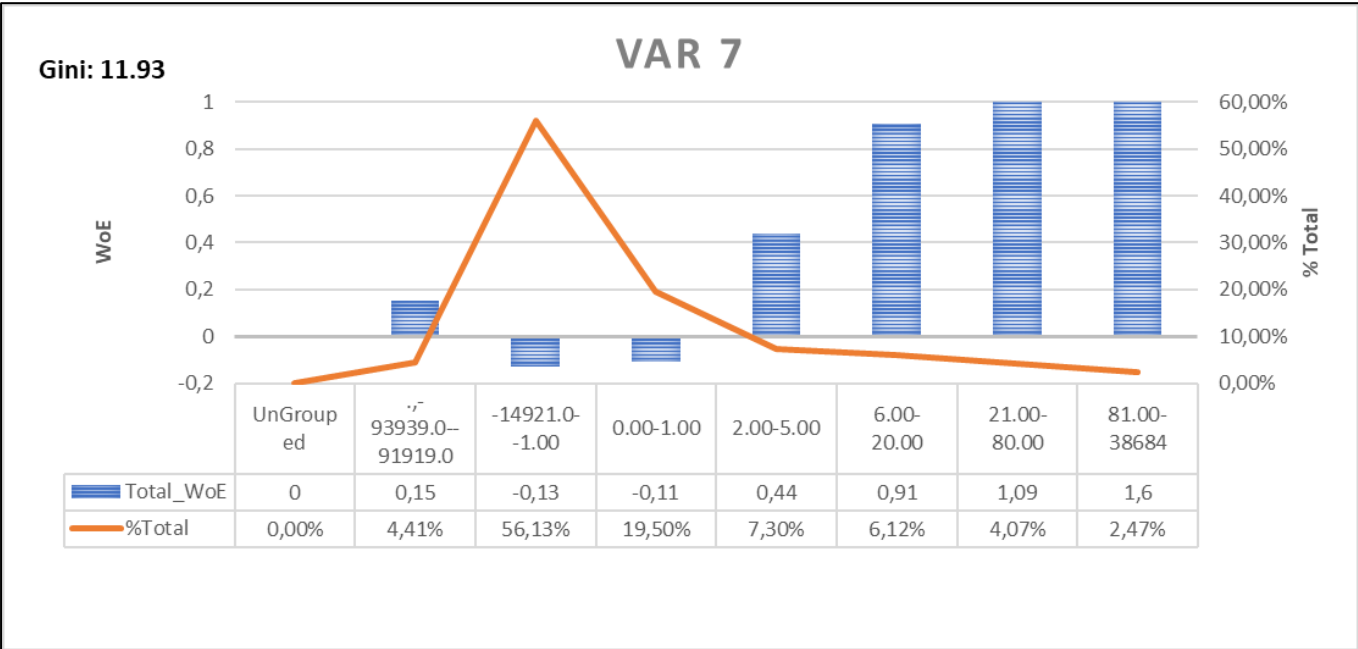


Figure 4. 11: WoE graph Var 7

The WoE in the above graph of Var 7 is first negative and then becomes positive indicating that the distribution of bads is higher between groups -14921.0-1.00 than in groups > 1.00. According to business logic the bad rate for this variable would decrease as the ratio increases, thus indicating the distribution of goods would be greater in subsequent groups. The ungrouped is not taken into consideration in the trend due to no volumes and group., -93939.0--91919.0 is not taken into consideration in the trend because they are known a special code.

The positive WoE in groups 2.00 – 38684 suggests a decrease in the risk of credit defaults thus increasing the predictive power in these groups. There is low predictive power in groups 14921.0-1.00. The variable has a monotonic relationship between the risk of defaulting and Var 7. Applications that fall in buckets 14921.0 – 1.00 are more likely to get a worse score as compared to the applications that would fall in buckets 2.00 – 38684.

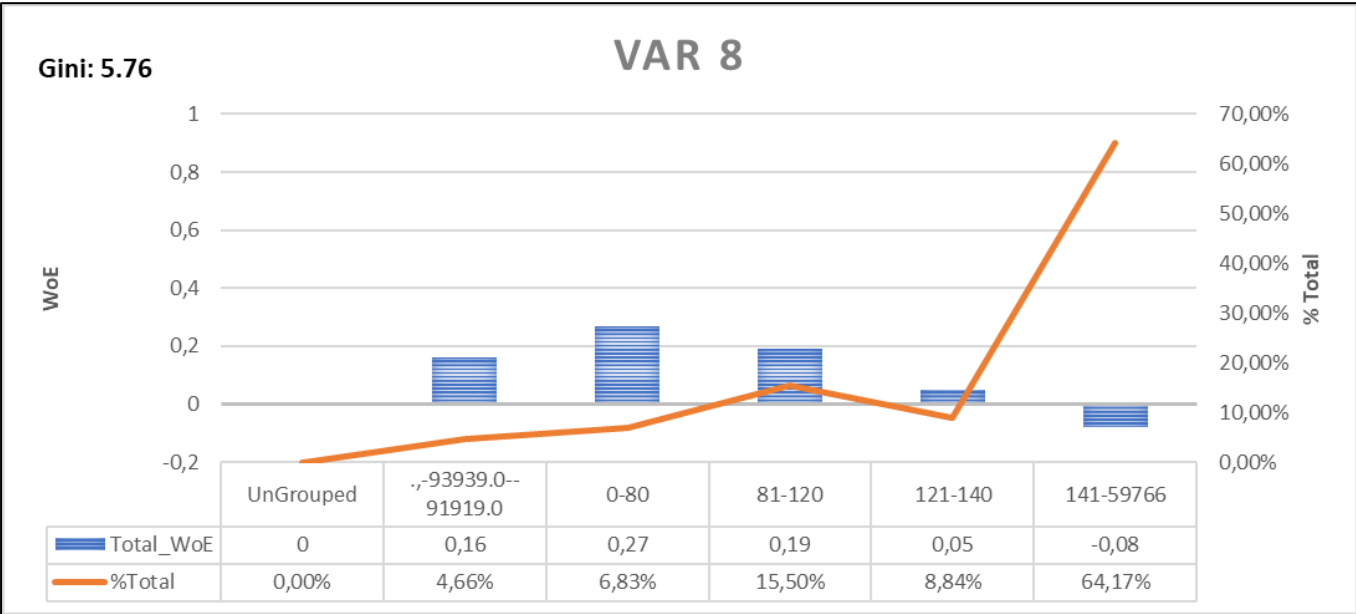


Figure 4. 12: WoE graph Var 8

The WoE graph in Figure 4.12 of Var 8 is first positive and then becomes negative indicating that the distribution of goods is higher between groups 0-120 than in groups > 120. According to business logic the bad rate for this variable would increase as the volatility increases, thus indicating the distribution of bads would be greater in subsequent groups. The ungrouped is not taken into consideration in the trend due to no volumes and group., -93939.0--91919.0 is not taken into consideration in the trend because they are known a special code.

The WoE in buckets 0-140 demonstrates that the risk of defaults occurring in these buckets are much less than in bucket 141-58766. The predictive power in buckets 0-140 is stronger thus suggesting that application would more likely get scored better in these buckets. A monotonic relationship can be observed thus exhibiting linearity across the WoE of the different buckets.

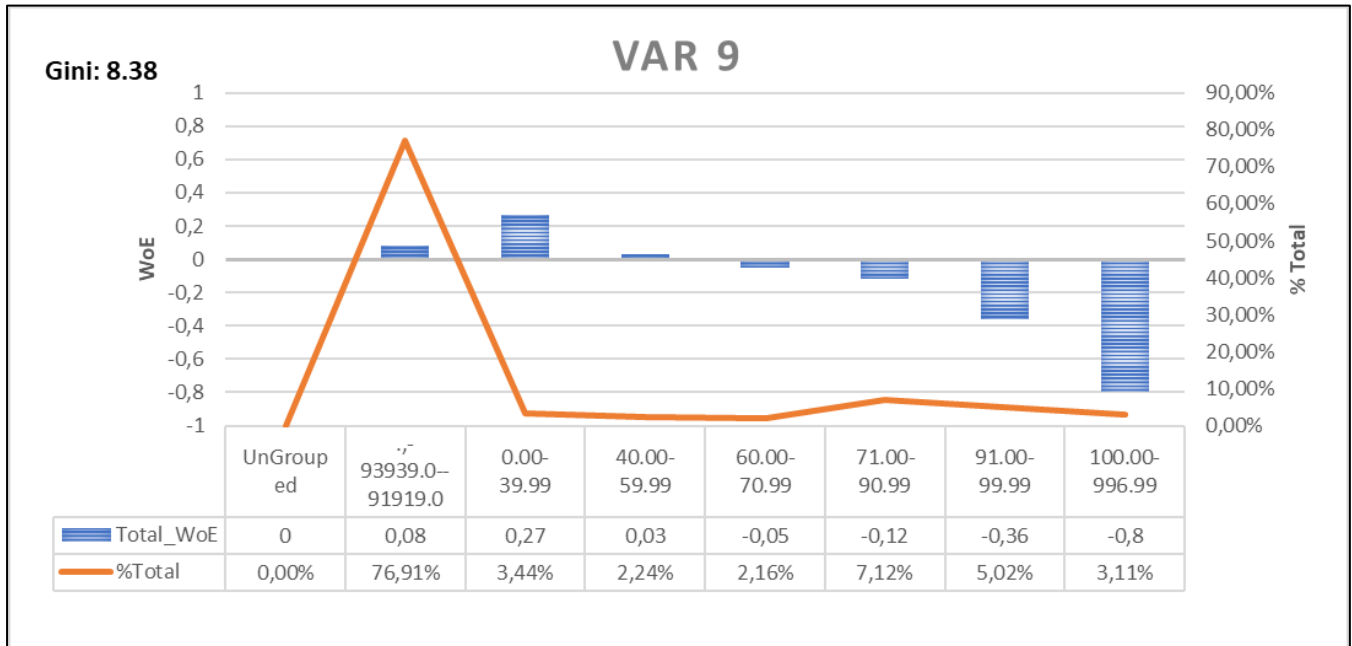


Figure 4. 13: WoE graph Var 9

The WoE in the graph above of Var 9 is first positive and then becomes negative indicating that the distribution of goods is higher between groups 0-59.99 than in groups > 59.99. According to business logic the bad rate for this variable would increase as the ratio increases, thus indicating the distribution of bads would be greater in subsequent groups. The ungrouped is not taken into consideration in the trend due to no volumes and group., -93939.0--91919.0 is not taken into consideration in the trend because they are known a special code.

Application that fall in buckets 60.00-996.99 would more likely get scored worse as the falling in these buckets would indicate the risk of defaulting to increase. There is a linear relationship across the buckets suggesting a monotonic relationship between the risk of defaults and the variable. In bucket 0.00-39.99 the predictive power is stronger which suggests the a decrease in defaults and customers to score better in this bucket than the others.

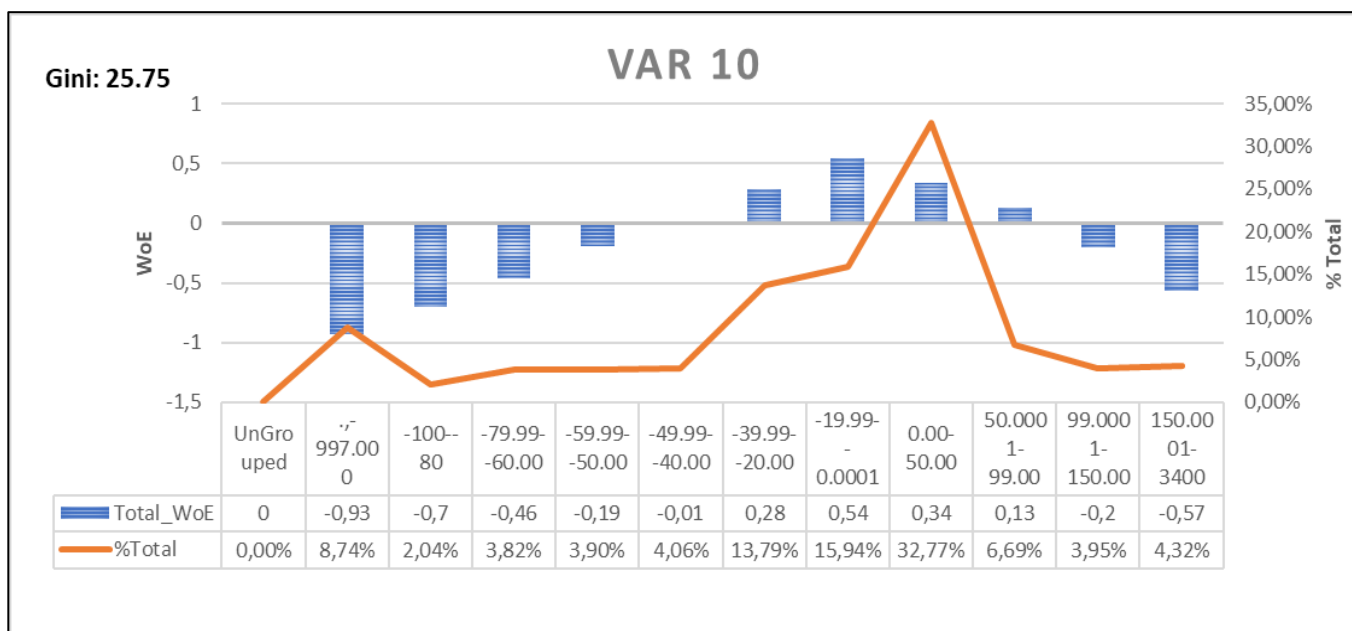


Figure 4. 14: WoE graph Var 10

The WoE graph in Figure 4.14 of Var 10 is first negative, then it becomes positive and then negative again. Indicating that distributions of bads is high for groups -100 - -50, then decreases for groups -39.99 – 99.00 and increases again for groups > 99.00. According to business logic the bad rate for this variable is parabolic, the bad rate would increase then decrease and begin to increase again. There is a plateau between groups -49.99 - -40.00 indication the distribution of goods and bads are equal. The ungrouped is not taken into consideration in the trend due to no volumes and group., -997.00 is not taken into consideration in the trend because they are known a special code.

The negative WoE in buckets -100 - -50.00 and buckets 99.000 – 3400 indicates the buckets where applications would score worse if they fall in these buckets as the probability of defaulting in these buckets are higher. This shows that the predicting an outcome in these buckets may be more difficult as the predictive power will be lower in these buckets. Buckets -39.99 – 99.00 are the buckets where applications would score better as the probability of defaulting in these buckets are lower. Predicting an outcome in buckets -39.99 – 99.00 would be less difficult as the predictive in these buckets are stronger and thus defaulting will be less likely to occur.

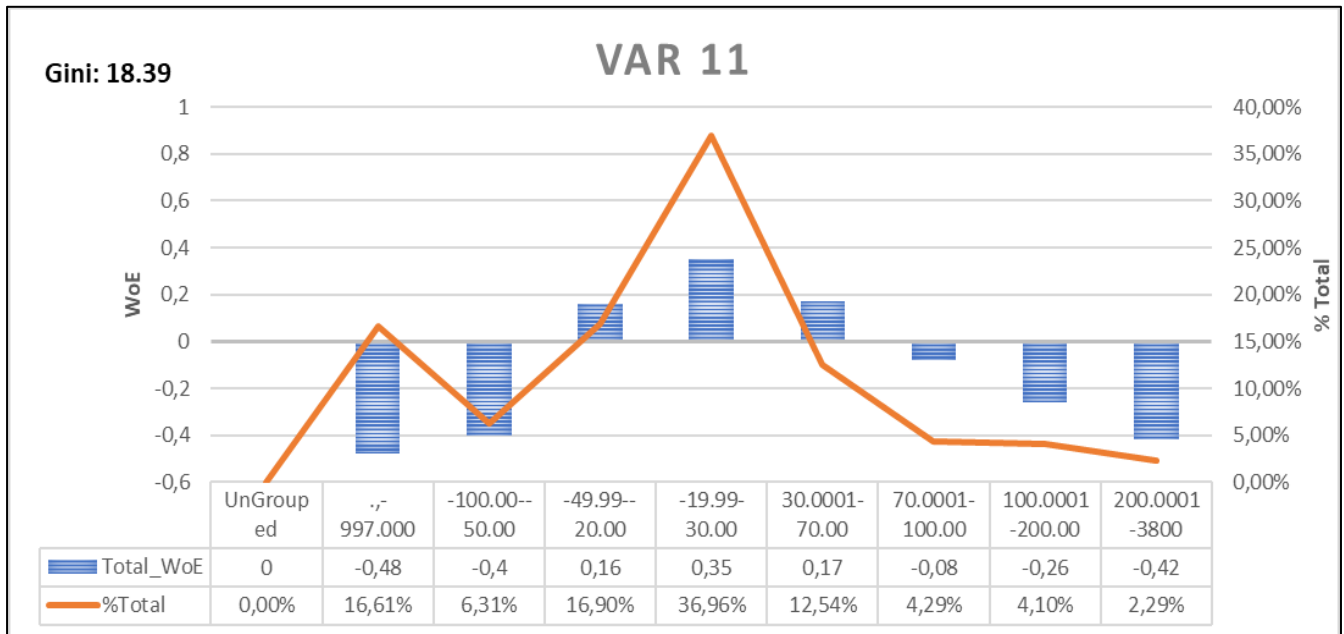


Figure 4. 15: WoE graph Var 11

The WoE graph in Figure 4.15 of Var 11 is first negative, then it becomes positive and then negative again. Indicating that distributions of bads is high for groups -100 - 50, then decreases for groups -49.99 – 70.00 and increases again for groups > 70.00. According to business logic the bad rate for this variable is parabolic, the bad rate would increase then decrease and begin to increase again. There is a plateau between groups -49.99 - -40.00 indication the distribution of goods and bads are equal. The ungrouped is not taken into consideration in the trend due to no volumes and group., -997.00 is not taken into consideration in the trend because they are known a special code.

Buckets -49.99–70.00 have a higher predictive power thus the probability of defaulting in these buckets will be less likely to occur, however in buckets -100.00-50.00 and buckets 70.0001-3800 have a lower predictive power thus the probability of defaulting in these buckets are more likely to occur.

The monotonic relationship exists between the risk of credit defaults and Var 11, there is a linear relationship between the WoE values in the Var 7.

4.10 Stability over time analysis

Stability Over Time Analysis: Logistic regression

Stability over time analysis or better known as the stability trend is done to see how the variable performs over time. It gives an indication of if a shift as occurred or not and gives an indication of the magnitude of the shift.

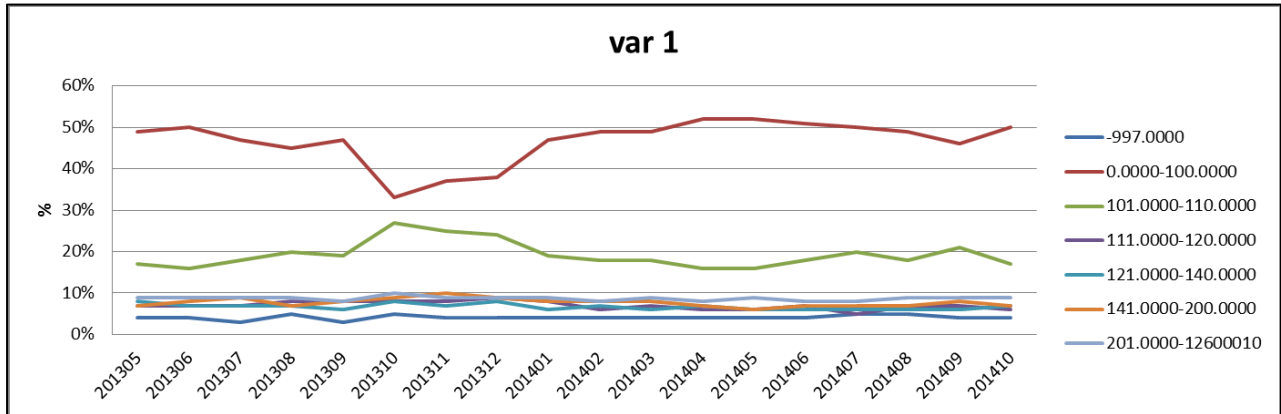


Figure 4. 16: Stability over time graph Var 1

Variable 1 is stable and there seems to be sudden spikes between 201310 – 201311. This can be due to economic change or a seasonal cycle such as school holidays or due to low volumes; otherwise the variable is stable over time.

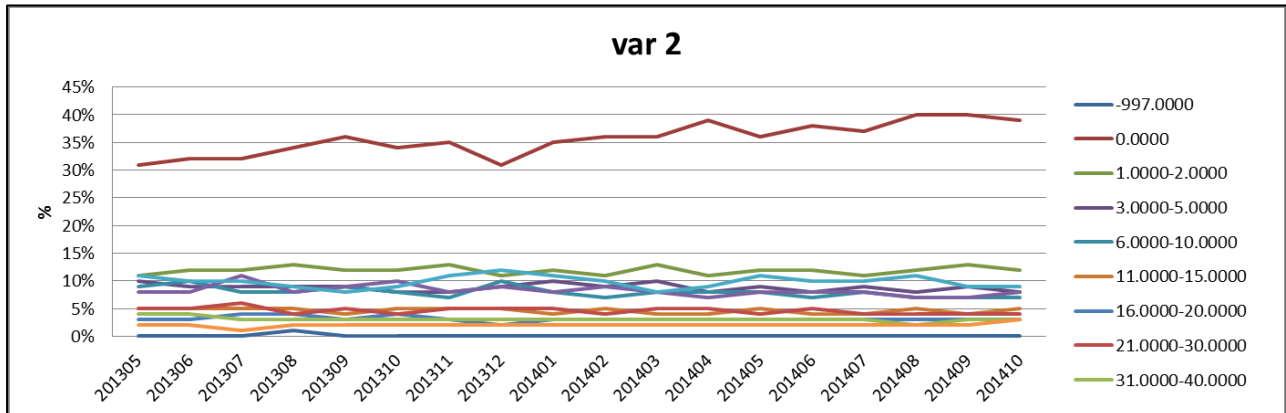


Figure 4. 17: Stability over time graph Var 2

Variable 2 is stable there are no major spikes present, suggesting that the variable is stable and there are no major shifts.

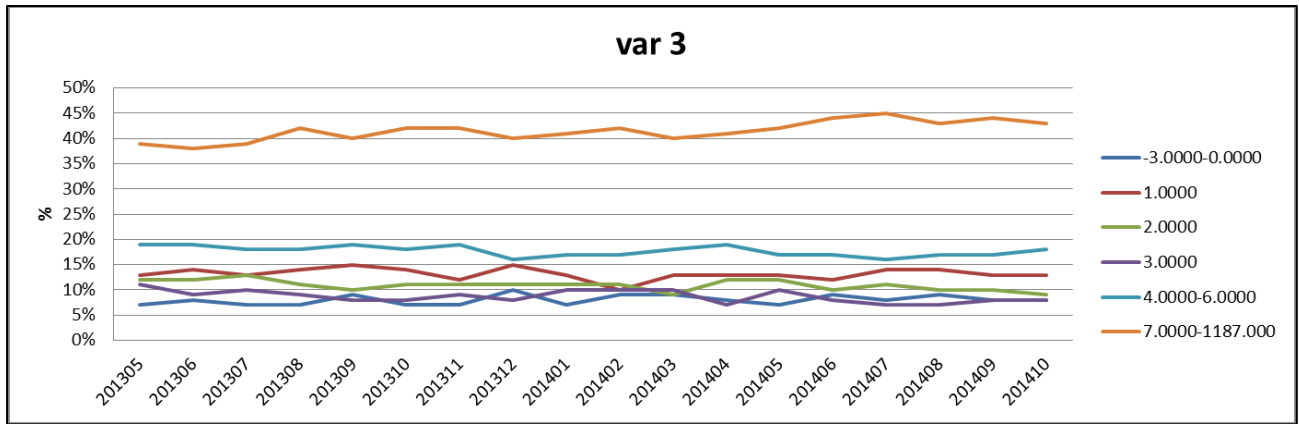


Figure 4. 18: Stability over time graph Var 3

Variable 3 is stable, there are no major spikes present, suggesting that the variable is stable and there are no major shifts.

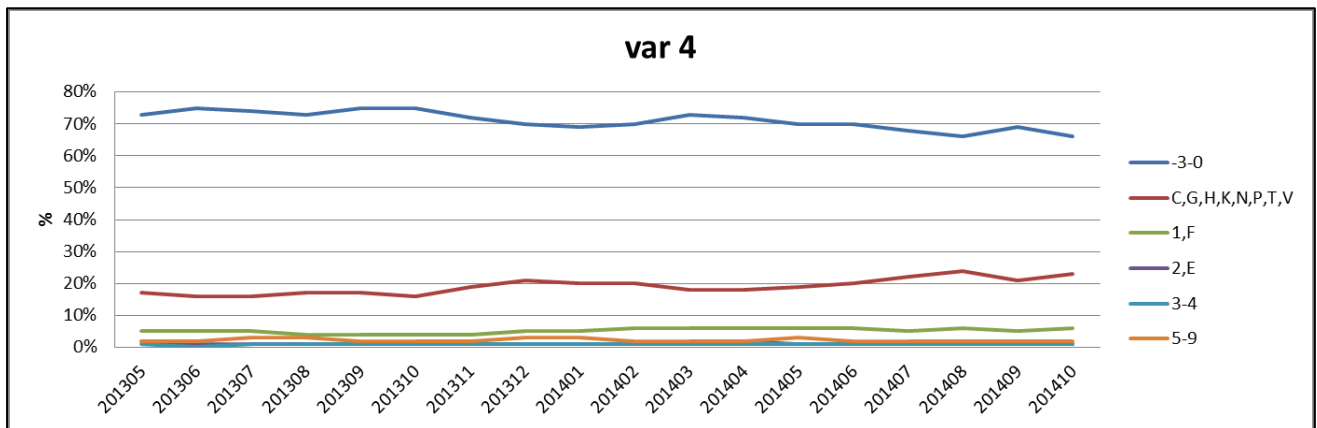


Figure 4. 19: Stability over time graph Var 4

Variable 4 is stable, there are no major spikes present, suggesting that the variable is stable and there are no major shifts.

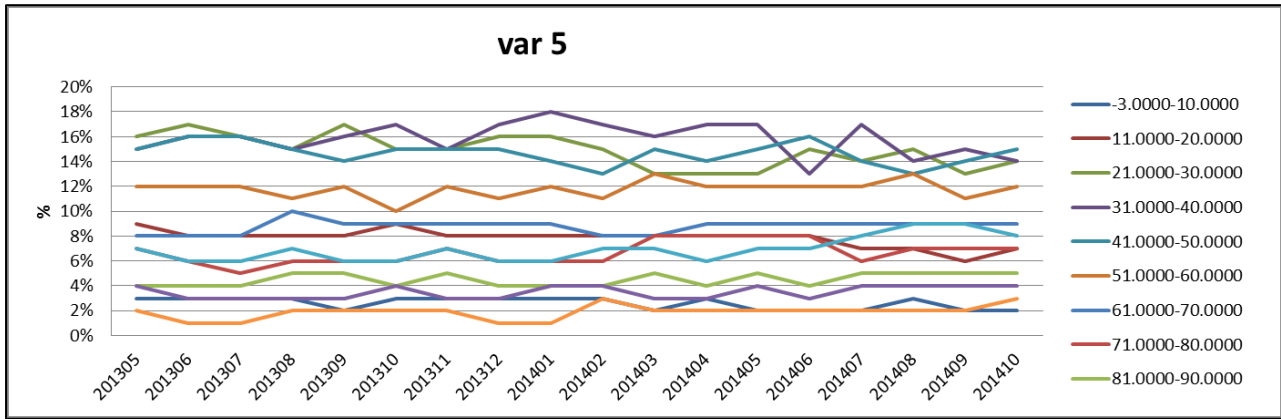


Figure 4. 20: Stability over time graph Var 5

Variable 5 is stable, there are no major spikes present. The small spike present at 201406 and 201407 can be due to economic change or a seasonal cycle such as school holidays or due to low volumes; otherwise the variable is stable over time, indicating no major shifts.

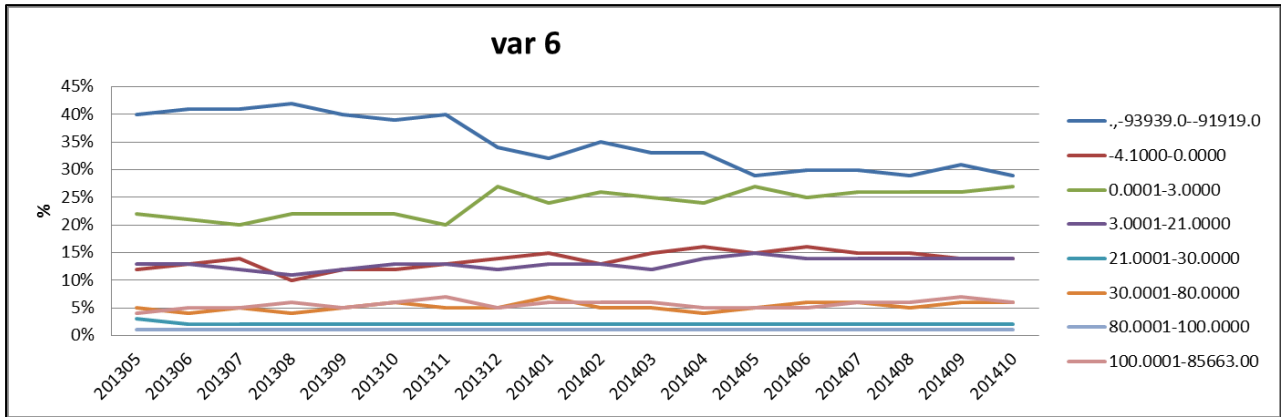


Figure 4. 21: Stability over time graph Var 6

Variable 6 is stable, there are no major spikes present, suggesting that the variable is stable and there are no major shifts.

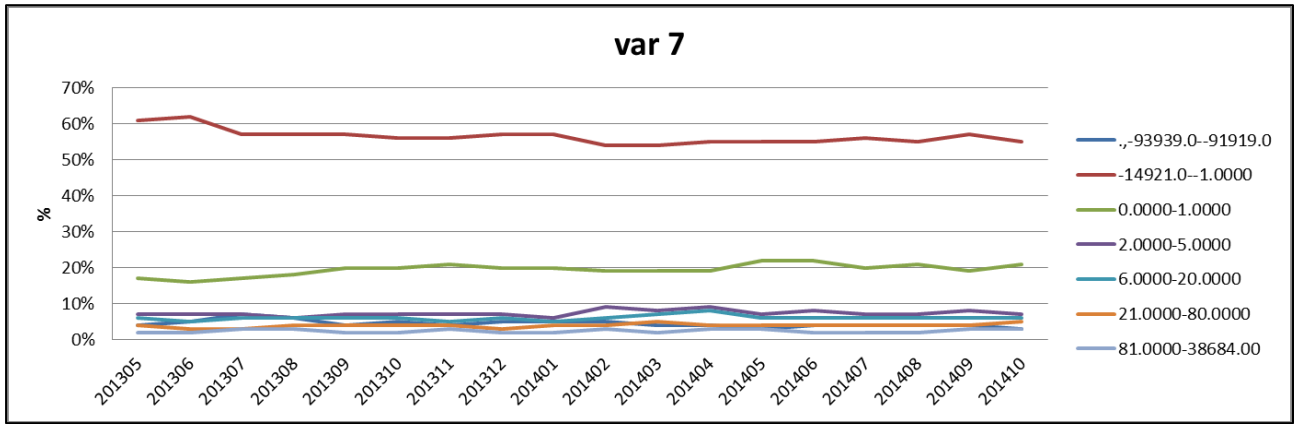


Figure 4. 22: Stability over time graph Var 7

Variable 7 is stable there are no major spikes present, suggesting that the variable is stable and there are no major shifts.

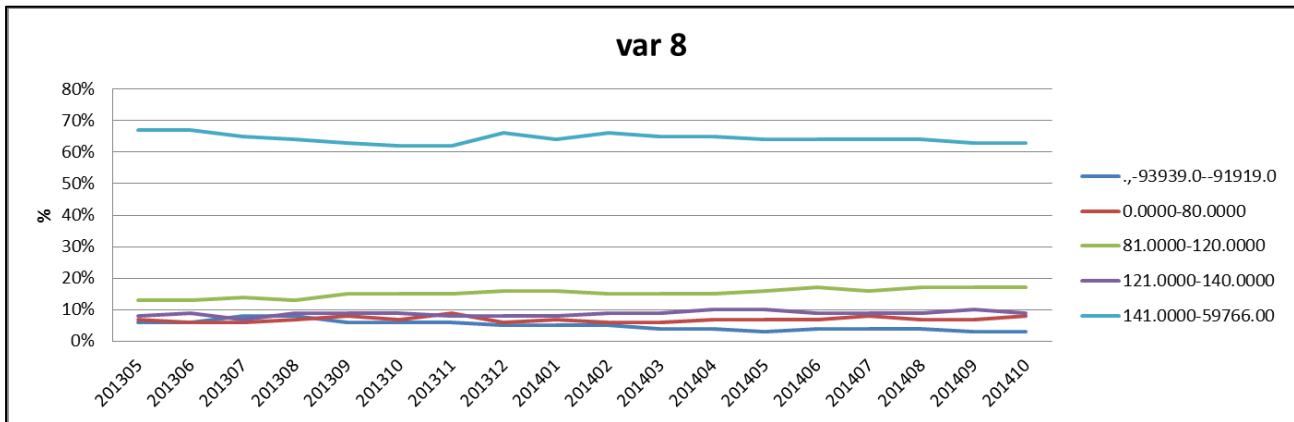


Figure 4. 23: Stability over time graph Var 8

Variable 8 is stable, there are no major spikes present, suggesting that the variable is stable and there are no major shifts.

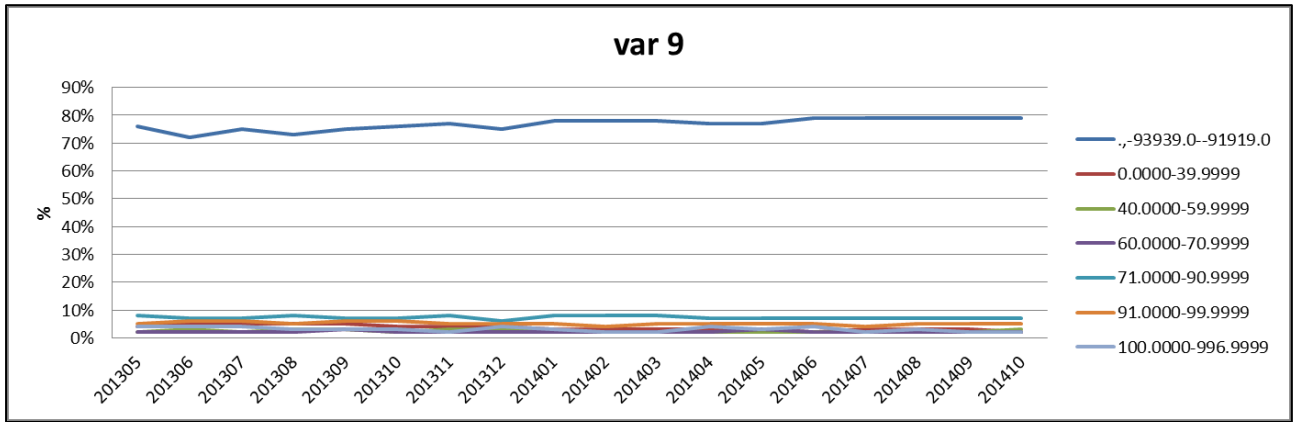


Figure 4. 24: Stability over time graph Var 9

Variable 9 is stable, there are no major spikes present, suggesting that the variable is stable and there are no major shifts.

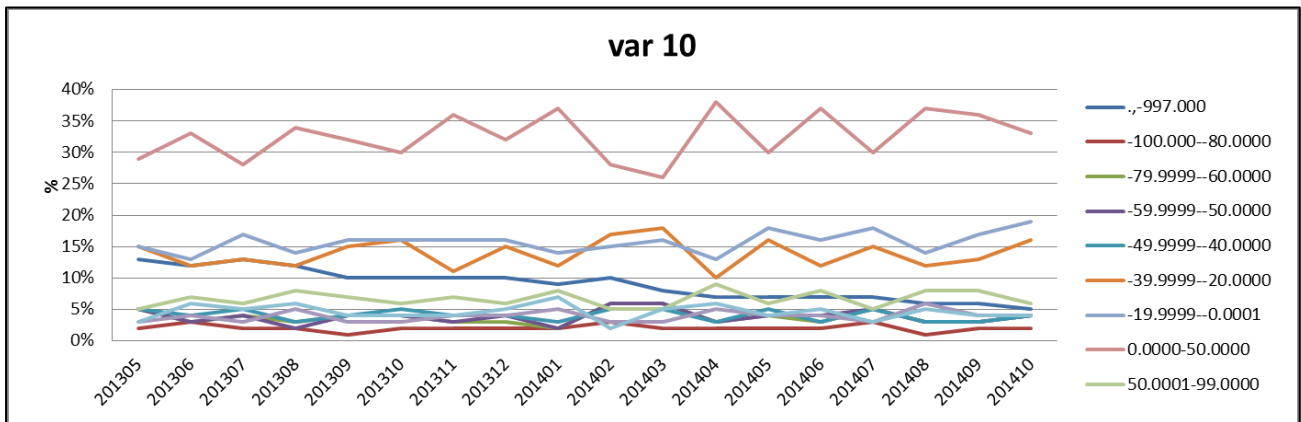


Figure 4. 25: Stability over time graph Var 10

Variable 10 is stable, there are no major spikes present. The small spikes present can be due to economic change or a seasonal cycle such as school holidays, or due to low volumes; otherwise the variable is stable over time.

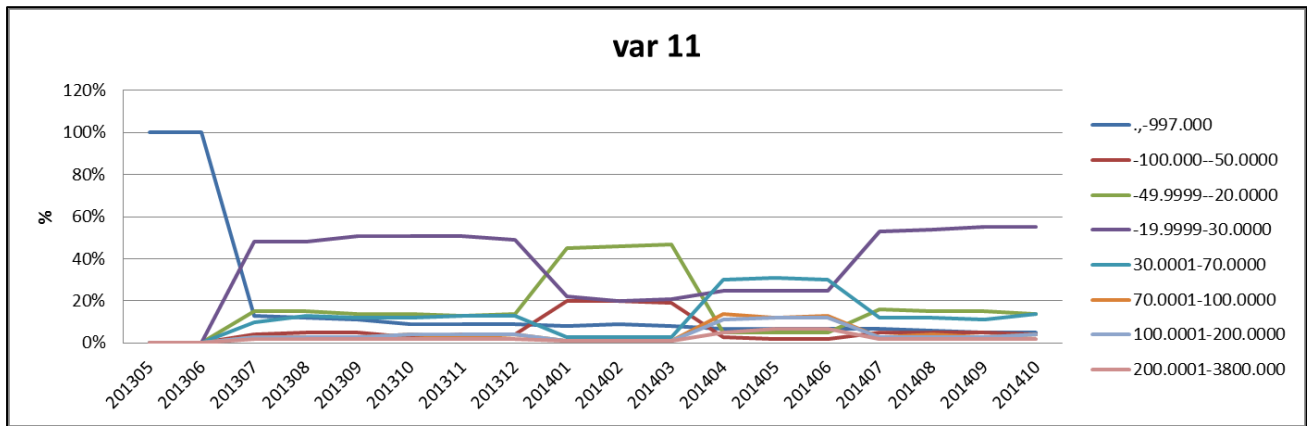


Figure 4. 26: Stability over time graph Var 11

Variable 10 is stable; there is an unusual trend present due to the fact that it is a seasonal variable. Trends tend to be a bit unpredictable because of economic changes; however, it does become stable over time.

4.11 Summary

A logistic regression model was used to build an application scorecard, from the results the proposed model has a fair, stable model, with a strong enough Gini to improve prediction accuracy with the coefficients without having to sacrifice any of its desirable features. The model produced by using the logistic regression is sufficient.

CHAPTER 5: CLASSIFICATION AND REGRESSION TREES

The following chapter will discuss classification and regression trees (CART) and the results obtained when using CART to develop a scorecard. When using CART there are classification and decision factors that need to be considered, CART then makes use of recursive partitioning during the tree building process that is of a binary nature due to the dependent variable being of a dichotomous nature. The next step is the pruning, which is done to make the tree smaller, pruning helps to prevent overfitting the model. Once the tree is built various results are analysed such as the Gini, the cumulative lift, classification tables, fit statistics, assessment plot and leaf statistics plot.

5.1 Introduction

Classification and Regression Trees (CART) falls under the machine learning methodology and is used in predictive modelling. CART involves binary recursive splitting of data in which the response variable is categorical and the “class” where the response variable falls into is classified by the tree; whereas in a regression tree the response variable is continuous, and a value is predicted as opposed to a classification being done. The structure of the CART algorithm is a sequence of questions, the next question would be determined by the answer given from the previous question.

5.2 Classification and decision factors

CART is comprised of four fundamental elements. The first element is the dependent variable generally known as the outcome variable; which is the variable being predicted based on the independent variable. The second element is the independent variable also known as the predictor variable; these variables are usually associated with the dependent variable. The third element is the “learning” dataset which consists of values for the dependent and independent variables. The final element is the “test” dataset or more commonly known as a validation dataset, which is used to verify the decision rules for splitting the tree.

5.3 Binary recursive partitioning

Recursive partitioning is described by selecting the split at each step of the algorithm such that the reduction in the residual sum of squares is maximized. The residual sum of squares is given by (Mason.gmu.edu, n.d.):

$$RSS = \sum_{m=1}^M \sum_{i \in R_m} (y_i - \hat{y}_{R_m})^2 \quad (5.1)$$

Since the above approach is tedious to compute the RBS (recursive binary splitting) approach is thus used. RBS starts at the top of the tree and splits the tree into two branches creating a partition of two spaces. This continues with this specific split from the top of the tree numerous times and selects the split with the features that minimises the current residual sum of squares. The tree constructs a new branch in a specific partition and continues the same procedure, evaluating the residual sum of squares of the partition and selecting the most optimal.

5.4 Pruning

Pruning is a method used in machine learning to decrease the size of decision trees by deleting segments of the tree that contribute little power to classify cases. When decision trees are constructed several of the branches may reveal the outliers of the training dataset; thus, pruning is used to assist with overfitting.

Post-Pruning

Post pruning is conducted after the tree is constructed, since the tree might be overfitted. Commonly known pruning strategies are known, namely:

1. Minimum Error – the tree is trimmed back to a stage where the cross-validated error is at a minimum. Cross-validation is a method used to test the efficiency of machine learning algorithms.
2. Smallest tree – the tree is trimmed back marginally further than the minimum error. A decision tree is created with the cross-validation error in 1 standard error of the minimum error.
3. Cost-Complexity Pruning – is error-based pruning which focuses on loss minimisation and probability estimation instead of error minimisation.
4. Reduced Error Pruning – test data is used to construct a series of pruned trees as opposed to using it only for the best tree selection.

5. Pessimistic Pruning – does not need a test data set; when using a training data set to construct a tree the misclassification rate is excessively confident, when used for pruning excessively large trees are constructed.
6. Critical Value Pruning – during the construction of the original tree, a goodness of split measure is used to determine at very node the associated attribute. The value of measure is used to demonstrate how well the attribute splits the data amongst the classes at that node. A critical value is specified and nodes that do not reach this critical value are pruned except if there is another node along the branch that does not meet the critical value. The pruning will be greater is a large critical value is chosen thus resulting in a smaller tree.

When comparing these methods, the following should be taken into account, minimum-error pruning is very reactive to the number of classes in the data. Significantly different levels of pruning even on the same data set. It can be seen as the least accurate method as it also is unsuccessful in providing a set of trees to be analysed.

The quickest method is the Pessimistic pruning which is the most “crude”, it does not require a separate test data set. This method can yield bad results and when being used should be handled with caution.

The Critical Value pruning and Reduced Error pruning methods generally produce low error rates. They require a test data set to able to produce a set of trees to analyse. When using the Critical Value pruning method one of the requirements needed is to specify the initial and incremental values this can be slow if they are not selected properly, however if it is used concurrently with a probability measure, actual significance levels can be determined. Further studies need to be done to investigate the interaction between the type of measure used in tree construction and the pruning method.

5.5 Classification and regression tree

When using the CART analysis; the assessment measure used is decision, with the interval target criterion set to variance and the nominal target criterion set to Gini. When building the tree, the GBI_IND (Good, Bad, Indeterminate Indicator) was used as the target where 1 is classified as bad and 0 is classified as good. The data set was spilt 80:20; 80% training (development) and 20% validation (holdout).

The Gini of the CART model built was 44.5

5.5.1 Variables with Gini

The list of variables with their Gini's shows the predictive power of each variable that was selected to be in the tree.

Gini	Variable	Description
21,60	Var 1	Payment Profile Var 1
18,70	Var 2	Payment Profile Var 2
14,99	Var 3	Payment Profile Var 3
22,38	Var 4	Payment Profile Var 4
18,29	Var 5	Payment Profile Var 5
9,07	Var 6	Behavioural Var 1
10,30	Var 7	Behavioural Var 2
17,80	Var 8	Debt Trajectory Var 1
20,48	Var 9	Payment Profile Var 6
12,37	Var 10	Transactional Var 1
14,19	Var 11	Behavioural Var 3
16,08	Var 12	Affordability Var 1
15,22	Var 13	Debt Trajectory Var 2
12,03	Var 14	Affordability Var 2
18,15	Var 15	Payment Profile Var 7
9,79	Var 16	Transactional Var 2
18,72	Var 17	Payment Profile Var 8
10,47	Var 18	Behavioural Var 4
16,62	Var 19	Payment Profile Var 9
11,10	Var 20	Debt Trajectory Var 3
12,31	Var 21	Debt Trajectory Var 4
7,68	Var 22	Payment Profile Var 10
11,53	Var 23	Debt Trajectory Var 5
9,74	Var 24	Behavioural Var 5
17,81	Var 25	Debt Trajectory Var 6
11,00	Var 26	Payment Profile Var 11
10,58	Var 27	Debt Trajectory Var 7
16,21	Var 28	Behavioural Var 6

Gini	Variable	Description
13,69	Var 29	Payment Profile Var 12
15,57	Var 30	Debt Trajectory Var 8
14,52	Var 31	Debt Trajectory Var 9
9,61	Var 32	Payment Profile Var 13
10,47	Var 33	Payment Profile Var 14
8,63	Var 34	Payment Profile Var 15
17,20	Var 35	Payment Profile Var 16
15,54	Var 36	Payment Profile Var 17
12,24	Var 37	Payment Profile Var 18

Table 5. 1: Classification and Regression Tree model variables

5.5.2 Importance of the variables

Variable importance is used to indicate which of the predictors are most useful for predicting the response variable. The table below shows the variable importance of the training set and the validation set as well as the variable importance of the ratio of the validation to the training set.

Variable	Training Importance	Validation Importance	Ratio of Validation to Training Importance
Var 1	1,00	1,00	1,00
Var 2	0,83	0,78	0,94
Var 3	0,72	0,64	0,88
Var 4	0,65	0,63	0,97
Var 5	0,52	0,41	0,78
Var 6	0,52	0,24	0,46
Var 7	0,50	0,31	0,62
Var 8	0,41	0,46	1,11
Var 9	0,41	0,00	0,00
Var 10	0,37	0,00	0,00
Var 11	0,36	0,40	1,13
Var 12	0,36	0,30	0,85
Var 13	0,34	0,35	1,04

Variable	Training Importance	Validation Importance	Ratio of Validation to Training Importance
Var 14	0,32	0,24	0,73
Var 15	0,32	0,34	1,08
Var 16	0,31	0,23	0,75
Var 17	0,28	0,38	1,35
Var 18	0,26	0,23	0,89
Var 19	0,23	0,26	1,12
Var 20	0,23	0,17	0,74
Var 21	0,23	0,14	0,61
Var 22	0,22	0,00	0,00
Var 23	0,22	0,00	0,00
Var 24	0,18	0,16	0,85
Var 25	0,18	0,22	1,21
Var 26	0,17	0,00	0,00
Var 27	0,16	0,00	0,00
Var 28	0,15	0,00	0,00
Var 29	0,15	0,16	1,09
Var 30	0,15	0,09	0,62
Var 31	0,14	0,00	0,00
Var 32	0,14	0,08	0,57
Var 33	0,14	0,00	0,00
Var 34	0,14	0,00	0,00
Var 35	0,13	0,21	1,60
Var 36	0,12	0,00	0,00
Var 37	0,10	0,00	0,00

Table 5. 2: Classification and Regression Tree model variable importance list

5.5.3 Cumulative lift

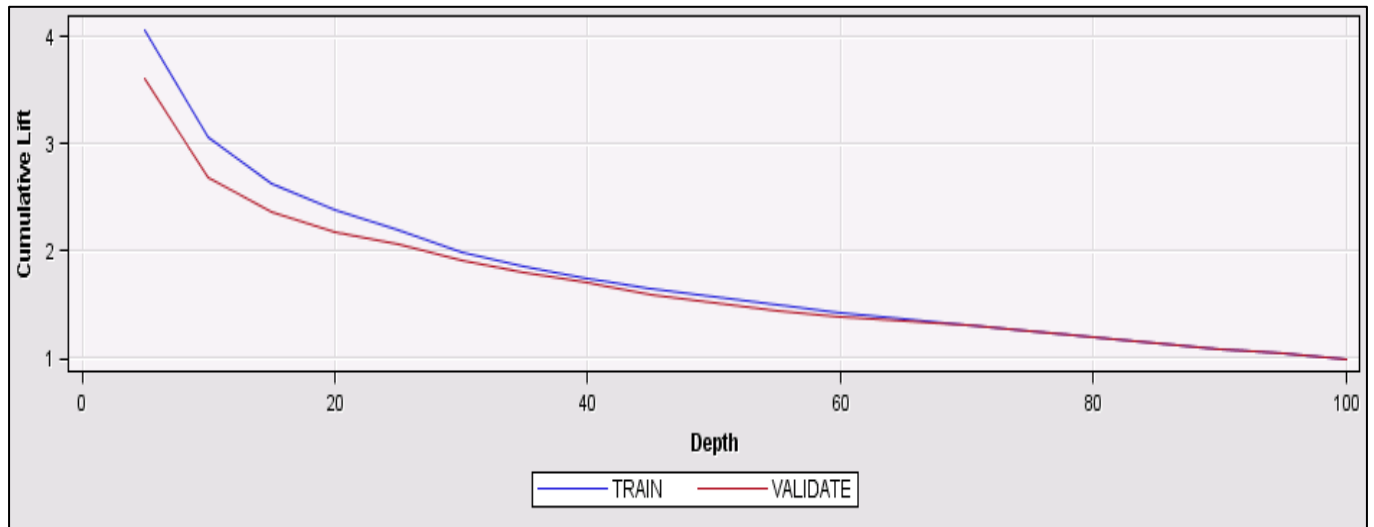


Figure 5. 1: Cumulative lift of Classification and Regression Tree model

The cumulative lift is the cumulative ratio of the percentage of targets up to the decile of interest to the percentage of targets in the entire data set. For the cumulative lift, the higher value in the lower deciles indicates a predictive scorecard model.

To explain how effective the model is the cumulative lift is used. The x-axis shows the percentile and the y-axis shows lift. If a random population of 10% is selected without using a model, we would get 10% of the at-risk customers by default; this is what is meant by no lift (or lift = 1). The chart shows that using the model that is given, one would be able to capture 32% - 35% of the at-risk customers; which is 3 times more than using no model, this is the lift over the baseline.

If we move on the lift curve to the right and select the top 20%, the lift is about 2.5% - 2.6%; which means that when using the model, 50% of the at-risk customers are captured.

5.5.4 Classification tables

Training Data Set

Event Classification Table					
Role=TRAIN		Variable=GBI_IND			
Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	96,16	98,14	128405,34	93,96
1	0	3,84	88,17	5130,00	3,75
0	1	78,00	1,86	2439,16	1,78
1	1	22,00	11,83	688,00	0,50

Table 5. 3: Classification table for training data set in Classification and Regression Tree model

The above table shows that for the training data set 93.96% of the population the target was good and the outcome was good (true positive); 3.75% of the population the target was bad but the outcome was good (false negative); 1.78% of the population the target was good but the outcome was bad (false positive), and 0.50% of the population the target was bad and the outcome was bad (true negative).

Validation Data Set

Event Classification Table					
Role=VALIDATE		Variable=GBI_IND			
Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	96,11	97,90	32454,92	93,78
1	0	3,89	90,18	1313,00	3,79
0	1	82,95	2,10	695,88	2,01
1	1	17,05	9,82	143,00	0,41

Table 5. 4: Classification table for validation data set in Classification and Regression Tree model

The above table shows that for the validation data set 93.78% of the population target was good and the outcome was good (true positive); 3.79% of the population target was bad but the outcome was good (false negative); 2.01% of the population target was good but the

outcome was bad (false positive), and 0.41% of the population target was bad and the outcome was bad (true negative).

5.5.5 Fit statistics

Fit Statistics	Train	Validation
Sum of Frequencies	25411	6356
Misclassification Rate	0,22	0,23
Maximum Absolute Error	0,95	1
Sum of Squared Errors	8052,75	2081,71
Average Squared Error	0,16	0,16
Root Average Squared Error	0,4	0,4
Divisor for ASE	50822	12712
Total Degrees of Freedom	25411	.

Table 5. 5: Fit statistics of Classification and Regression Tree model

The fit statistics are used to measure the goodness of fit of a statistical model and tells us how well the model fits the set of observations. The fit statistics usually summarizes the inconsistency between the expected values and observed values under the model in question.

The **misclassification rate** shows the percentage classified instances that are correct during the classification. The calculation indicates the percentage of records that the model has placed in the incorrect group or class. The formula is given as follows:

$$\text{Misclassification rate} = \frac{\text{total no. of misclassification}}{\text{total no. records classified}} \quad (5.2)$$

The misclassification rate for the training data set is 22%; indicating that only 22% is misclassified and the rest of the 78% of the training data set is not. The misclassification rate for the validation data set is slightly higher at 23%; indicating that only 23% is misclassified and the rest of the 77% of the validation data set is not.

The **sum of squared errors (SSE)** is the differences between each observation and its group mean. It may be used as a measure of variation in a cluster; if all instances in a cluster are identical then the sum of squared errors will be zero.

Average Squared Error is the average squared difference between the estimated values and what is estimated. An averaged squared error value that is closer to zero indicates that the model has a reduced random error component and the fit will be more beneficial for prediction. The average squared error in the model for the training and validation data set are both 0.16; which is considerable close to zero indicating a reduced random error component.

The Root Average Squared Error also known as the Root Mean Square Error (RMSE) is known as the standard deviation of the residual (prediction of the errors). The measure of how far from the regression line data points lie are known as the residuals, the measure of the distance of the residuals from the regression line is known as the root mean square error (RMSE); therefore the RMSE gives an indication of the concentration of the data around the line of best fit.

The model has an RMSE of 0.4 indicating that the model somewhat does not improve prediction over the model drastically; since the value of the RSME ranges between 0-1, where 0 indicates that the proposed model does not improves prediction over the mean model and 1 indicating perfect prediction. The value in our model is 0.4 which is slightly close to zero but close to halfway to 1, which shows a slight but not great improvement of prediction.

5.5.6 The assessment plot

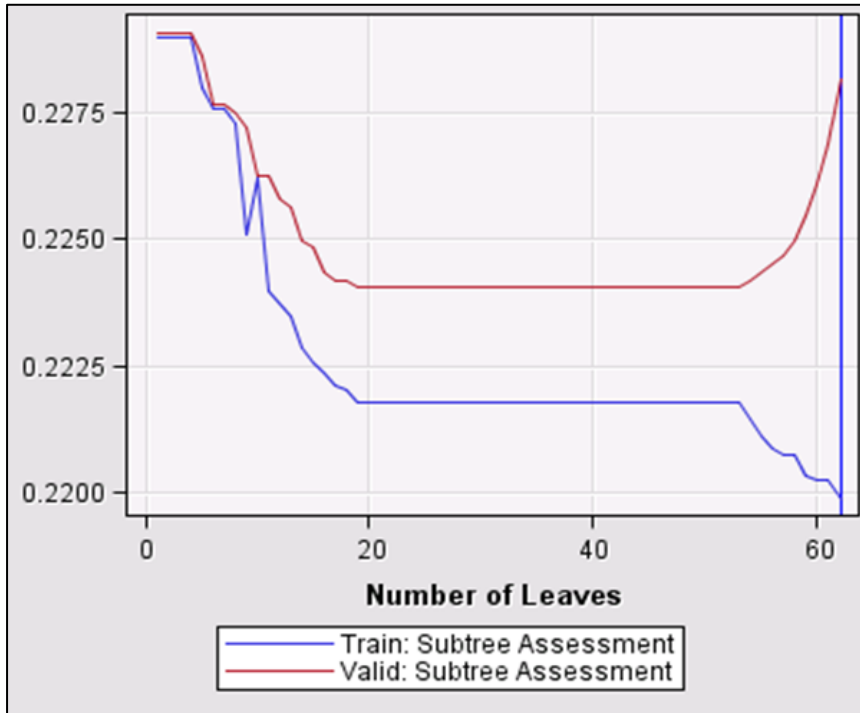


Figure 5. 2: Assessment plot of Classification and Regression Tree model

The Assessment Plot shows tree evaluation information. Trees are evaluated using the proportion of cases that are correctly classified in the top 50% of the training data. For each tree size, a tree that correctly classifies the maximum training cases is chosen to represent that size. The chosen tree is evaluated once more with validation cases. The following conditions should be identified:

- If the line of the training data is progressing
- The extent to which the validation data confirm the progress
- How much progress seems achievable while using the smallest tree possible.

The line for the training data is progressing as the number of leaves increases, validation data confirm this progress, but starting from the tree that has 12 leaves and 60 leaves, there is a slight difference.

5.5.7 The leaf statistics plot

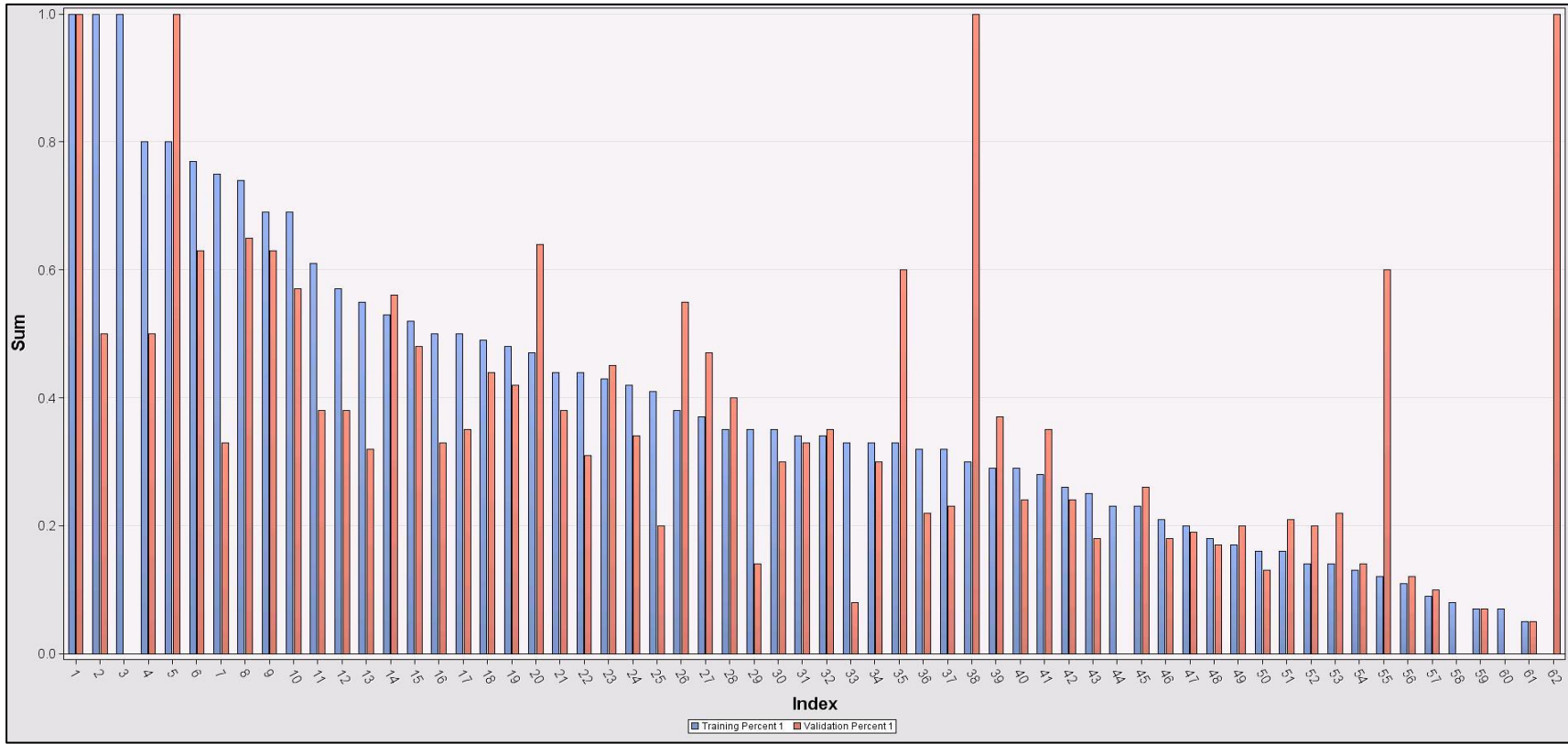


Figure 5. 3: The leaf plot of the Classification and Regression Tree model

The bars represent the proportion of events ($BAD=1$) for each leaf for the training data set (blue bars) and the validation data set (red bars). The bars are sorted in order of increasing values of the proportion of events in the training data. If the tree is reliable, a similar increase in bar heights will be seen in the validation data, except in some of the small leaves where there is not enough data to make reliable estimates. Overall, you do see this increase, but some of the leaves look unreliable; this is because the proportion of events is higher in the training data than in the validation data and vice versa: the proportion of events is higher in the validation data than in the training data.

5.6 Summary

The CART method produced a model with a Gini of 44.5%. the model was stable and produced the most “purest” tree as possible. As the number of leaves on the tree increases the better the classification becomes according to the assessment plot. The misclassification rate for the validation data set is 23%; indicating that only 23% is misclassified and the rest of the 77% of the validation data set is not. Thus, the CART method produced a stable, adequate model.

CHAPTER 6: SUMMARY, CONCLUSION AND RECOMMENDATIONS

6.1 Introduction

The present chapter will be used to summarize the chapters in the study, the results obtained from the study and recommendations of the research.

6.2 Chapter 1: Summary

6.2.1 Introduction

Chapter one gives an overview of the study by discussing the problem statement, objectives and methodology.

6.2.2 Problem statement

Financial institutes need to distinguish between good and bad customers when granting credit. One of the most common methods used is credit scoring, which calculates an associated risk score to each applicant. Two modelling techniques used in this study to create a credit scoring model is the logistic regression and classification and regression trees (CART). Each with their own pros and cons, the aim of the study to identify which technique is the most efficient.

6.2.3 Objectives of the study

The following objectives were formulated for the study:

6.2.3.1 Primary objectives

Two different scorecards were built using different statistical modelling techniques, namely logistic regression and classification and regression trees.

6.2.3.2 Theoretical objectives

The primary objectives were achieved by formulating the following theoretical objectives for the study: (a) understanding what credit scoring entails and (b) understanding the two modelling techniques being compared.

6.2.3.3 Empirical objectives

In accordance with the primary objective of the study, the following empirical objectives were formulated: two scorecards were developed using two different modelling techniques and then the results were compared. The aim was to find an effective modelling technique that would have a stronger and more accurate predictive power.

6.2.4 Research design and methodology

6.2.4.1 Empirical study

The scorecard was built for Overdrafts and Revolving Loans facility for a financial institute. The model was built for existing customers applying for an Overdraft or Revolving Loan facility or for a limit increase on a current Overdraft or Revolving Loan facility. The scorecard will determine the risk at application stage of applicants applying for an Overdraft or Revolving Loan and/or limit increase.

6.2.4.2 Statistical analysis

Logistic Regression Approach

The scorecard that was developed using the logistic regression technique was built using Paragon's scorecard modelling tool, Modeller, under license from Paragon. The modelling approach used was a Logistic Weights of Evidence (WoE) regression modelling with stepwise elimination.

CART Approach

The scorecard that was developed using the CART technique was built using SAS Enterprise Miner, under license from SAS.

6.2.5 Ethical considerations

All the information that was used regarding the description of the variables was kept discrete and thus renamed due to the sensitivity of the data used. Authorisation has been given to use actual variables provided they are renamed and used in an appropriate manner.

6.3 Chapter 2: Summary

6.3.1 Introduction

There are primary stages in the development and implementation process of a scorecard. The primary stages include Data Exploration and Data Cleaning, Characteristic Analysis, Preliminary Scorecard, Reject Inference, Final Scorecard and Validation.

6.3.2 Stage 1: Data Exploration

Data exploration involves the steps taken to collect the data and clean data. Characteristic analysis and variable grouping were done in data exploration.

6.3.3 Preliminary Scorecard: Known Good Bad Model (KGB)

The Known Good Bad model was developed on the population who have taken up (TU) or not taken up (NTU) a product. Therefore, the population was “known”, the model was built with a distribution that was adapted by using the previously classified accept/reject weights

6.3.4 Reject Inference

Reject inference was done to give a more realistic view of the performance outcome of the entire population. It is used to see what the performance would of the previously rejected applications be if these applications were accepted.

6.3.5 Final Scorecard: All Good Bad model (AGB)

The All Good Bad model was the final model built after reject inference had been done. Once the final model was built the scaling of scores, points allocated, misclassification and scorecard strength needed to be finalised.

6.3.6 Validation

Validation of the model entails verifying whether the performance of the model will still be acceptable on a different data set.

6.3.7 Post-Implementation

Once the scorecard was validated, implementation needed to begin. Once implemented on the systems, the scorecard needed to be monitored and reports needed to be produced to make sure the scorecard is still stable and performing as expected.

6.4 Chapter 4: Summary

6.4.1 Introduction

The chapter was about the scorecard that was built with logistic regression and results.

6.4.2 Assumptions of a logistic regression

The dependent variable assumes that the event in question had either occurred or did not occur. There are no outliers present, this was achieved by converting the continuous predictors to uniform scores. There should be no correlation present between the predictors, that is, no correlation of more than 0.5 is allowed. The logistic regression depends on analytically estimated coefficients, Euler's constant and the log-odds transformation to predict probability.

6.4.3 Logistic regression results

When using the logistic regression method, stepwise selection is used with scaled odds, scaling does not influence the predictive strength of the scorecard.

In a stepwise regression, predictor variables that do not statistically improve the model are not kept in the model; when the best fitted model is reached the log-odds of the likelihood is calculated by the model. However, there are instances when this does not yield the most appropriate model.

6.4.4 The Gini coefficient

The Gini coefficient or Gini is a statistic to measure the predictive power of the scorecard and as well as the variables in the scorecard. If the Gini is 0% then the scorecard cannot distinguish between good and bad applicants.

6.4.5 Known Good Bad (KGB) model

The Known Good Bad model was developed on the "known" population, that is, the population the has taken up or not taken up a product. The results were validated.

6.4.6 Accept Reject (AR) model

The Accept Reject model was developed on the population that was accepted and declined following the same thought pattern as for the KGB. This model was predicting the accept/reject and not good/bad.

6.4.7 Reject Inference

Reject inference was performed once the KGB and AR models were developed and validated. Reject inference was done to determine what would happen to the rejected applications.

6.4.8 All Good Bad (AGB) model

Once the KGB, AR and reject inference was done the final model known as the AGB was developed and validated. Scaling of scores, points allocated, misclassification and scorecard strength was finalised and completed.

6.4.9 Weights of Evidence trend

The value of the variables used in the model build are not the characteristic values themselves, but the Weight of Evidence of the group to which it belongs.

6.4.10 Stability over time analysis

Stability over time analysis was done to determine if each variable is stable over time and that are no shifts that occur in the variable over time.

6.5 Chapter 5: Summary

6.5.1 Introduction

Classification and regression trees are one of many machine learning techniques used in credit scoring. The response variable is binary and recursive splitting is done on the data to form a tree like structure.

6.5.2 Classification and decision problems

Four essential factors are embodied by CART. These factors help test and validate the model and to verify the rules to split the tree.

6.5.3 Binary and recursive partitioning

Recursive partitioning was done to select the best split at each step such that the reduction in the residual sum of squares is maximized.

6.5.4 Pruning

Pruning is the technique used to decrease the size of the decision trees; this is achieved by deleting segments of the tree that do not contribute to classifying cases. Post-pruning is done after the tree is constructed.

6.5.5 Classification and regression tree

When the scorecard was developed using the CART technique, the assessment measure that was used was decision, with the interval target criterion set to variance and the nominal target criterion set to Gini. The chosen variables with their Gini was recorded. Analysis on the variable importance, cumulative lift, classification tables, fit statistics, the assessment plot and the statistics plot were validated and completed.

6.6 Results

The first scorecard was built using a logistic regression. The model was built in paragon modeller. The process followed when using the logistic regression technique was done by first completing a data analysis then segmentation is done. Once segmentation is finalized the KGB is built and validate, followed by the AR with validation done. Once the KGB and AR model were built, reject inference was then applied, thereafter once the reject inference was validated the AGB model was built and validated.

The second scorecard was built using classification and regression trees, the model was built in SAS eMiner. The data analysis was done and then segmentation was completed. The trees were constructed and pruned where necessary.

The Logistic Regression technique built a final model with a Gini of 56 whereas the CART technique built a final model with a Gini of 44.5. The Gini is a statistic used to measure the ranking ability of the scorecard as well as the variables in the scorecard, in other others, the predictive power of the scorecard and variables in the scorecard. Even though the Logistic Regression built a model with a higher Gini than the CART technique, the CART technique produces other statistics which suggest the technique has built a stable and robust model. The misclassification rate of the CART model was 23% whereas in the logistic regression model the holdout sample and the development sample do not deviate from each other.

To improve the accuracy of prediction of a tree and the accuracy of its splits is a “balancing act”. Searching for splits may result in selection bias which may allow pointless variables to be selected; however, avoid overlooking decent splits that may be hidden behind interactions or linear relationships. Significance tests can be done to help solve this problem by effects increasing in complexity and priority of the order decreasing. Tests for mains effects are always conducted; tests for interactions are performed only if no main effect test is significant; lastly tests for linear structure are done only if no test in the previously mentioned tests is significant. A Bonferroni correction influences the significance level of every group. Furthermore, if an interaction test is significant, the split is obtained by a two-level search on the set of relating variables.

6.7 Conclusion and recommendations

In this paper we compared and assessed the logistic regression model results with the CART model results for credit scoring. The evaluation was done by observing the performance of the classification accuracy and the complexity of the variables in each model. It was observed that the logistic regression model outperformed the CART model. The logistic regression model had a better performance and by using the proposed methodology, robust and simple models were achieved for credit scoring.

Further work can be conducted on different and larger data sets and applying the proposed methodology to other classification methods. Other work that can be done, is to show the impact of using significance tests for the splitting rule in the CART technique.

BIBLIOGRAPHY

ABDOU, H. A. & POINTON, J. 2011a. Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management*, 18, 59-88.

ABDOU, H. A. & POINTON, J. 2011b. Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management*, 18, 1-41.

AGRESTI, A. 2002. Inference for contingency tables. *Categorical Data Analysis, Second Edition*, 70-114.

ANDERSON, R. 2007. *The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation*, Oxford University Press.

ANONYMOUS Classification and Regression Tree (CART) Techniques. Winter School on "Data Mining Techniques and Tools for Knowledge Discovery in Agricultural Datasets", 12.

ANONYMOUS Classification: Basic Concepts, Decision Trees and Model Evaluation.

ANONYMOUS Logistic Regression. NCSS NCSS Statistical Software.

ANONYMOUS 2010. Logistic Regression.

ANONYMOUS 2011. Classification and Regression Tree (CART) Techniques. 222-233.

ANONYMOUS 2013. *CreditScoring Guide A step-by-step guide to building application scorecards using CreditScoring by TrimTab*. TrimTab: TrimTab Solutions.

BARBON, F. & Vidal, M.F. 2019. How to use advanced analytics to build credit-scoring models that increase access. NW, MSN IS7-700, Washington, DC 20433 USA, 1-40.

BERGERUD, W. 1996. Introduction to Regression Models: with worked forestry examples. Biom. Info. Hand. 7. Res. Br., BC Min. For., Victoria, BC Work. Pap, 26, 1996.

BHALLA, D. (2019). *ListenData*. [online] ListenData. Available at: <https://www.listendata.com/> [Accessed 10 Feb. 2020].

CALLEJÓN-FERRE, A.J., CARREÑO-ORTEGA, A., DÍAZ-PÉREZ, M. & SALINAS-ANDÚJAR, J. A. 2019. Application of Logistic Regression Models for the Marketability of Cucumber Cultivars. *Journal of Agronomy*, 9,17,1-19.

CAMDEVIREN, H. A., YAZICI, A. C., AKKUS, Z., BUGDAYCI, R. & SUNGUR, M. A. 2007. Comparison of logistic regression model and classification tree: An application to postpartum depression data. *Expert Systems with Applications*, 32, 987-994.

CAPON, N. 1982. Credit scoring systems: A critical analysis. *The Journal of Marketing*, 82-91.

CAVANAUGH, J. E. 2012. 171: 290 model selection lecture II: the Akaike information criterion. Retrieved December, 21, 2016.

CSH (2015). *Scorecard development*.

COKLUK, O. 2010. Logistic Regression: Concept and Application. *Educational Sciences: Theory and Practice*, 10, 1397-1407.

DE'ATH, G. & FABRICIUS, K. E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81, 3178-3192.

DUPUY, J.F., MONBET, V., POTERIE, A. & ROUVIÈRE, L. 2019. Classification tree algorithm for grouped variables, 1-39.

ECKEL, S. (2008). *Lecture 14: Interpreting logistic regression models*.

EVANS, D. L., DREW, J. H. & LEEMIS, L. M. 2008. The distribution of the Kolmogorov–Smirnov, Cramer–von Mises, and Anderson–Darling test statistics for exponential populations with estimated parameters. *Communications in Statistics—Simulation and Computation*®, 37, 1396-1421.

GREIFF, W. R. 1999. *Maximum entropy, weight of evidence and information retrieval*. University of Massachusetts Amherst.

GURNÝ, P. & GURNÝ, M. 2013. Comparison of Credit Scoring models on probability of default estimation for US Banks. *Prague economic papers*, 2, 163-181.

HOSMER JR, D. W., LEMESHOW, S. & STURDIVANT, R. X. 2013a. *Applied logistic regression*, John Wiley & Sons.

HUANG, E. & SCOTT, C. 2007. *Credit Risk Scorecard Design, Validation and User Acceptance*.

HYEOUN-AE, P. 2013. *An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain*. Korean Society of Nursing Science, 43, 11.

ITOO, R. A. & SELVARASU, A. 2015. RESEARCH ARTICLE: CASE STUDY OF CONSUMER CREDIT SCORING: A PROPOSAL. *International Journal of Latest Trends in Finance and Economic Sciences*, 6, 3627 - 3641.

IZENMAN, A. J. 2008. Recursive Partitioning and Tree-Based Methods. *Modern Multivariate Statistical Techniques*, 34.

KADANE, J. B. & LAZAR, N. A. 2004. Methods and criteria for model selection. *Journal of the American statistical Association*, 99, 279-290.

RASA KANAPICKIENE, R. & SPICAS, R. 2019. *Credit Risk Assessment Model for Small and*

Micro-Enterprises: The Case of Lithuania. *Journal risks*, 7, 1-23.

KAYRI, M. K. A. İ. 2015. The Comparison of Gini and Twoing Algorithms in Terms of Predictive Ability and Misclassification Cost in Data Mining: An Empirical Study. *International Journal of Computer Trends and Technology (IJCTT)*, 27, 10.

KLEIN, D. G. K. M. 2010. *Logistic Regression. A Self-Learning Text*, Springer Science Business Media.

KOON, S. and PETSCHER, Y. (2015). *Comparing Methodologies for Developing an Early Warning System*.

LEE, T.-S., CHIU, C.-C., CHOU, Y.-C. & LU, C.-J. 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, 50, 1113-1130.

LEUNG, K., CHEONG, F., CHEONG, C., O'FARRELL, S. & TISSINGTON, R. Building a scorecard in practice. *Proceedings 7th International Conference on Computational Intelligence in Economics and Finance*, Taiwan, 2008.

LEWIS, R. J. 2008. An introduction to classification and regression tree (CART) analysis. Available at: Accessed April, 27, 2008.

LOH, W.-Y. 2009. Improving the precision of classification trees. *The Annals of Applied Statistics*, 1710-1737.

LOUZADA, F., ARA, A. & FERNANDES, G. B. 2016. Classification methods applied to credit scoring: A systematic review and overall comparison. arXiv preprint arXiv:1602.02137, 1-50

LUCAS, A. (2004). *The Gini Coefficient*. [online] Rhinorisk.com. Available at: <http://rhinorisk.com/> [Accessed 24 Jul. 2019].

MAIMON, L. R. A. O. DECISION TREES. DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK.

MARQUEZ, J. 2008. An introduction to credit scoring for small and medium size enterprises. World Bank, Washington, DC.

MASON.GMU.EDU. (n.d.). *Chapter 8: Classification and Regression Trees*. [online] Available at: http://mason.gmu.edu/~jgentle/GT/L12_Tree_Methods.pdf [Accessed 1 Aug. 2019].

MESTER, L. J. 1997. What's the point of credit scoring? *Business review*, 3, 3-16.

MEYERS 2005. Chapter 3A: Data Screening. *Applied Multivariate Research*, 31.

MONTRICHARD, D. 2007. Reject Inference Methodologies in Credit Risk Modeling. Presentation in CIBIC.

PAEK, I. 2009. Three Statistical Testing Procedures in Logistic Regression: Their Performance in Differential Item Functioning (DIF) Investigation. ETS Research Report Series, 2009.

PARAGON BUSINESS SOLUTIONS (n.d.). *How MODELLER Calculates Score with Weight of Evidence Models*. pp.1-3.

PARAGON BUSINESS SOLUTIONS (2014). *Paragon Modeller Overview v1.2*. London: Paragon Business Solutions Limited.

RAWLINGS, J. O., PANTULA, S. G. & DICKEY, D. A. 1998. *Applied regression analysis: a research tool*, Springer Science & Business Media.

REZAC, M. & REZAC, F. 2011. How to measure the quality of credit scoring models. *Finance a Uver*, 61, 486.

SIAMI, M. & HAJIMOHAMMADI, Z. 2013. Credit scoring in banks and financial institutions via data mining techniques: A literature review. *Journal of AI and Data Mining*, 1, 119-129.

SIDDIQI, N. 2006. Credit risk scorecards. Developing and implementing intelligent credit scoring. SAS Institute Inc. New Jersey. USA, 1-196.

THOMAS, L. C., EDELMAN, D. B. & CROOK, J. N. 2002. Credit scoring and its applications, Siam.

TIMOFEEV, R. 2004. Classification and regression trees (CART) theory and applications. Humboldt University, Berlin.

VOJTEK, M. & KOÈENDA, E. 2006. Credit-scoring methods. Czech Journal of Economics and Finance (Finance a uver), 56, 152-167.

WALTER W. HAUCK, J. A. A. D. 1977. Wald's Test as Applied to Hypotheses in Logit Analysis. Journal of the American Statistical Association, 72.

WU, X. 2008. Credit Scoring Model Validation. Faculty of Science, Korteweg-de Vries Institute for Mathematics.

YAP, B. W., RANI, K. A., RAHMAN, H. A. A., FONG, S., KHAIRUDIN, Z. & ABDULLAH, N. N. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013), 2014. Springer, 13-22.

YOHANNES, Y. & HODDINOTT, J. 1999. Classification and regression trees: an introduction. International Food Policy Research Institute, 2033.

YOHANNES, Y. & WEBB, P. 1999. Classification and regression trees, CART: a user manual for identifying indicators of vulnerability to famine and chronic food insecurity, Intl Food Policy Res Inst.

ZAMBON M, L. R., BUNN A, AND POWELL S 2006. Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis. Photogrammetric Engineering & Remote Sensing, 72, 25-30.

