2019

An Investigation into the Predictive Capability of Customer Spending in Modelling Mortgage Default

Donal Finn [Thesis]
Technological University Dublin.

Recommended Citation

Finn, D. (2019) An Investigation into the Predictive Capability of Customer Spending in Modelling Mortgage Default, Masters Thesis, Technological University Dublin.

An Investigation into the Predictive Capability of Customer Spending in Modelling Mortgage Default

Donal Finn

A dissertation submitted in partial fulfilment of the requirements of Technological University Dublin for the degree of M.Sc. in Computer Science (Data Analytics)

2019

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Data Analytics), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the test of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Technological University Dublin and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed: __________________________________

Date: **14 June 2019**

ABSTRACT

The mortgage arrears crisis in Ireland was and is among the most severe experienced on record and although there has been a decreasing trend in the number of mortgages in default in the past four years, it still continues to cause distress to borrowers and vulnerabilities to lenders. There are indications that one of the main factors associated with mortgage default is loan affordability, of which the level of disposable income is a driver. Additionally, guidelines set out by the European Central Bank instructed financial institutions to adopt measures to further reduce and prevent loans defaulting, including the implementation and identification of Early Warning Indicators (EWIs). Financial institutions currently adopt credit risk models in order to calculate the risk associated with customers. Therefore, this research observed a cohort of mortgage customers in Lender A over a 30-month period and utilised transactional features, explaining the use of disposable income, to expand on existing credit risk models and aid in the identification of EWIs for the mortgage portfolio. Over the course of the study three feature selection techniques were adopted, namely correlation-based analysis, random forest feature importance and decision tree feature importance. A number of transactional categories were identified including insurance spend, gambling spend, savings and the value of ATM withdrawals. Furthermore, it was found that the inclusion of transactional features in existing credit risk models statistically improved performance.

Key words: *mortgage default, non-performing exposure, credit risk, logistic regression, transactional features, feature selection*

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisor, Dr. Sarah Jane Delaney for all the help, guidance and suggestions she has provided throughout this project.

I would also like to thank my colleagues in Lender A who facilitated me with access to the necessary data and whose experience and insights were greatly appreciated.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

1 INTRODUCTION

1.1 Background

European banks are presently experiencing high levels of non-performing exposures (NPEs). As a result of capital constraints faced by banks with high NPE levels, there is a general consensus of the view that high NPE levels lead to a negative impact on bank lending to the economy. At both a macro-prudential and micro-prudential view, a consistent reduction of NPEs in financial institutions will be beneficial to the economy.

In financial institutions, the development of credit risk models is influential in identifying customers who are likely to default on their loan. According to the ECB, Ireland has been recognised as one of the poorest performing member states, and while there has been a decline in NPEs, the Irish banking sector is required to conform to ECB guidelines and adopt the advisory measures to further reduce and prevent NPEs. One such measure is the implementation and identification of Early Warning Indicators (EWIs). This paper will set out to expand on the credit risk models developed in Lender A by including a set of transactional features relating to customer spending which will aid in the identification of EWIs for the mortgage portfolio.

1.2 Research Project

The aim of this research project is to develop transactional features based on customers' spending habits and assess their usefulness in predicting mortgage customers that will default on their repayment obligations within Lender A.

1.3 Research Objectives

The predictive models to be built throughout the experiments will include derived transactional features sourced internally from Lender A's Enterprise Data Warehouse (EDW) as well as a set of features used in current credit risk models within Lender A.

The objectives of this research are:

- To review the literature on mortgage arrears and default trends

- To review the literature and best practices for credit scoring and predictive modelling
- Design and develop transactional features to be assessed for predicting mortgage arrears
- Design experiments to test the hypothesis
- Train a baseline model using features available in existing credit risk model in Lender A for comparison and evaluation of the models built during experimentation
- Compare sampling methods to overcome class imbalance
- Apply feature selection methods to the transactional features to identify most predictive
- Train predictive models including most important transactional features
- Assess the results from predictive models including transactional features compared to baseline model to evaluate if it would be beneficial to include transactional features in future models
- Determine what future research could be undertaken to expand on the project

1.4 Research Methodology and Analytical Approach

The research methodology utilised in this project is an empirical evaluation, which will involve investigating and experimenting with a number of derived historical transactional features, developed based on customers spending behaviour. Experiments will be developed to ascertain the power of these features for predicting mortgage customers that may default at some stage in the future. In order to determine the success of the experiments undertaken in this research, appropriate performance measures such as recall and average class accuracy will be used, and suitable statistical techniques will be applied.

The experimental research undertaken will incorporate two overlapping areas: feature selection and two class classification. Both of these are common techniques in data mining. Typically, data mining is used to identify patterns from pre-processed and transformed data that are not obvious and where the number of permutations and sequences cannot be easily examined.

In the case of this research, it would be difficult to observe every customer who has a mortgage with Lender A and ascertain what their spending behaviour is indicating with regards the performance of their mortgage repayments. Data mining allows for techniques to be applied to a non-standard dataset such as this to learn the patterns in customer spending that commonly result in their mortgage defaulting.

1.5 Scope & Limitations

The scope of this project is to build a predictive model for mortgage customers in Lender A which utilises derived transactional features based of customer spending patterns. The aim of the experiment is to determine how useful these transactional features are in terms of predicting mortgage default and evaluate their value if their included in industry standard credit risk models within Lender A.

The project will include data from a population of customers who opened a mortgage post 2013 in Lender A. Each customer will be observed at six observation points between October 2017 and March 2018 and transactional features will be developed for each customer over the 12 months preceding each observation point. Due to the magnitude of the data, a full year of observation points was not made available by Lender A. Therefore, the data may not fully represent changes in customer spending behaviours due to seasonality.

1.6 Document Outline

The remaining chapters of this research are arranged into the Literature Review, Experiment Design and Methodology, Implementation and Evaluation and a Conclusion.

Chapter 2 can be looked at in two parts. The first part documents the current literature available on the banking crisis, mortgage arrears and the impact they have had on banks. This part will also cover research on some of the factors which influence mortgage arrears. The second part covers the literature review in the field of data mining, including two class classification, class imbalance, variable selection methods and model evaluation methods.

Chapter 3 discusses the experiment design and research methodology in place to deal with the class imbalance issue and to attempt to improve on existing credit risk models through the addition of transactional features. Variable selection and performance measures will also be discussed in this chapter.

Chapter 4 details the experiments undertaken and evaluates the models developed through the use of appropriate statistical techniques.

Finally, chapter 5 concludes the research with an overview of the contributions made by this paper to the problem of predicting mortgage arrears. Areas of applicability and potential future research are also discussed.

2 LITERATURE REVIEW

2.1 Introduction

This chapter discusses the relevant literature in the field of credit risk and default prediction of mortgage customers in financial institutions. The first sub sections (2.2 – 2.4) relate to the areas of mortgages and credit risk, detailing the evolution of mortgage arrears in Ireland, the impact they have on financial institutions and the key factors influencing them. The different methods associated with credit risk modelling and how transactional data is utilised in modelling credit risk from mortgage customers is also discussed. These sub-sections also cover challenges surrounding the decisions financial institutions must make prior to developing credit risk models whilst also outlining recommendations, from the literature, that were made to strengthen the field. The literature of transactional features is reviewed, and it is noted that there is a shortage of research relating to how transactional factors effect credit risk models. The review concludes detailing a successful example of the application of transactional factors in credit risk models in the United States. Through the research, it is found that mortgage lending is a significant problem and mortgage arrears are largely dependent on disposable income and loan affordability. Furthermore, it is recommended by the European Central Bank that lenders implement methods of establishing early warning indicators in an attempt to reduce non-performing exposures.

The remaining sub sections (2.5 – 2.10) in this chapter review literature in the field of knowledge discovery and data mining with particular focus on predictive modelling. Both knowledge discovery and data mining are explained and illustrated using widely used approaches and frameworks such as Knowledge Discovery in Databases (KDD) and the Cross Industry Standard Process for Data Mining (CRISP-DM). A review of the predictive models used for Classification and the current methodologies used by banks enables an understanding of the frequently used feature selection methods, model evaluation methods and model performance measures used to build a predictive model to assess mortgage customers that are likely to default and how they may be improved. The issue of class imbalance is also discussed with methods on addressing the issue reviewed through the literature.

2.2 Context

2.2.1 The Banking Crisis

The recent global financial crisis was precursor to the largest spate of international banking crises seen since the Great Depression of the 1930s. Due to the substantial financial, economic and societal impacts of the 2007 crash, much of the literature in this area strives to examine and determine the causes of the crash. In a study of 58 economies examining the causes of the financial crisis of 2007/2008, Claessens et al. (2010) found that factors including asset price bubbles, rapid credit growth and current account imbalances spread the crisis. In a 2018 study commissioned by the Bank of England, Aikmen et al.[1] also identified these factors in addition to inadequate regulation and/or supervision of regulation, excessive funding and elevated household debt, loose monetary policy and a belief that banking institutions were too big to fall as contributory to the global crisis. In Ireland, which is the focus of this study, the banking crisis began in 2008 and was systemic by 2009 (Leaven & Valencia, 2013), the impacts of which are still being felt by financial institutions and customers alike. The "Honohan Report" of 2010 presented three root causes of the banking crisis in Ireland specifically, which are intrinsically linked to the global causes outlined above.

- A regulatory approach which was insufficiently challenging and too accommodating resulting in delayed corrective regulatory intervention
- An under-resourced approach to bank supervision
- An unwillingness for the Central Bank & Financial Services Authority of Ireland to react to the risk of a forthcoming crisis

Furthermore, it is noted through multiple reports, that while linked to and influenced by the global economic crisis, the Irish banking crisis was largely "home-grown" (Honohan, 2010, Regling & Watson, 2010, Nyberg, 2011). The Commission of Investigation of the Banking Sector (Nyberg, 2011) notes that "the problems causing the crisis as well as the scale of it were the result of domestic Irish decisions and actions". Regling and Watson (2010) expand that the fundamental cause of the collapse was a property bubble "compounded by exceptional concentrations of lending for purposes

[1] https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2018/would-macroprudential-regulation-have-prevented-the-last-crisis

related to property". For context, these exceptional concentrations of lending translate to domestic property-related lending increasing by circa €200Bn over the period 2002-2008, representing 80% of all credit growth in the same timeframe (Nyberg, 2011). Following the collapse of the Irish credit bubble, the impacts of this were several fold. The most significant impact in the context of this study was the steep decline of Irish property prices and the concurrent increase of mortgage arrears.

2.2.2 Mortgage Arrears

Arrears are defined by the Central Bank of Ireland as arising "on a mortgage loan account where a borrower has not made a full mortgage repayment, or only makes a partial mortgage repayment, in accordance with the original mortgage contract, by the scheduled due date"[2]. Prior to the collapse of the Irish banking system in 2008, mortgage arrears in Ireland were low enough for the Central Bank not to collect or publish regular data on them, with the first significant data only being published from 2010. Following the economic crash, however, property prices in Ireland fell sharply with a concomitant increase in mortgage arrears. Data obtained for this study from the CSO[3] shows that residential property prices reached a peak in Q2 2007, crashing heavily over the period 2007-2013. A decline in property prices of the order of 55% over this period is observed, alongside an increase in mortgage arrears. Specifically, over the period 2009 to 2013, the number of Principal Dwelling Houses (PDH) mortgages in default increased by over 270% from 26,000 to just under 100,000. (Connor & Flavin, 2015). Presently, 63,246 PDH and 20,579 Buy-to-Let (BTL) mortgages remain in arrears, meaning that borrowers are unable and unlikely to make full repayments on their original contract to lenders. The mortgage arrears crisis in Ireland was and is among the most severe experienced on record (McCann, 2017).

Although as shown in figure 2.1 below, the levels of mortgages in arrears has declined over the period 2014-2018, over ten years on from the onset of the crisis mortgage arrears nonperforming loans continue to cause distress to borrowers and vulnerabilities to lenders.

[2] https://www.centralbank.ie/docs/default-source/Regulation/consumer-protection/other-codes-of-conduct/24-gns-4-2-7-2013-ccma.pdf?sfvrsn=4

[3] https://www.cso.ie/en/releaseandpublications/ep/p-rppi/residentialpropertypriceindexmarch2019/)

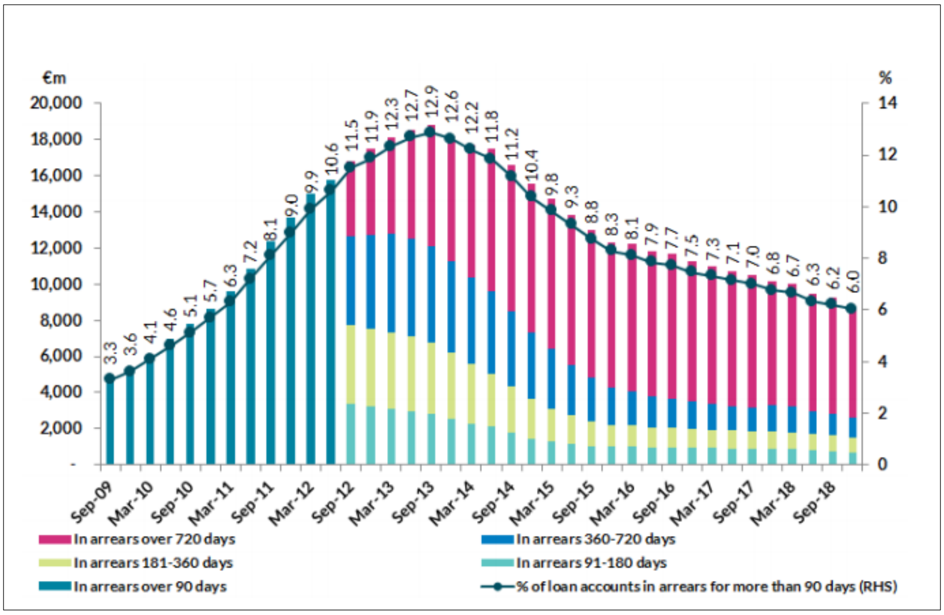


Figure 2-1: PDH Mortgage Accounts in Arrears over 90 Days

Significant problems remain and the financial impact should not be masked or underestimated by the downward trend shown above. The most recent quarterly report published by the Central Bank of Ireland[4], cite that of the 63,000 Principal Dwelling Houses in arrears, circa 44,000 mortgages were in arrears greater than 90 days as of December 2018. This translates to an outstanding balance of €8.7Bn on lenders, making up 8.8% of the outstanding balance of all PDH mortgaged accounts. Similar information may be derived from Central Bank data for Buy-To-Let (BTL) properties. As of December 2018, 15,600 BTL accounts were in arrears greater than 90 days, an outstanding balance of €4.3Bn and 22.3% of the outstanding balance of all BTL mortgages. Combined, this amounts to an outstanding balance of €12.9Bn on lenders, making up 12% of the outstanding balance of all mortgaged accounts. Including loans which are less than 90 days in arrears, the outstanding balance on lenders for all home loans in arrears amounts to €16.2Bn.

2.2.3 Impact of Arrears & NPEs on Banking Profitability

Considering PDH and BTL mortgages are the largest asset classes within Irish banks, the increase in arrears and NPEs between 2008 and 2013 had significant implications on

[4] https://www.centralbank.ie/docs/default-source/statistics/data-and-analysis/credit-and-banking-statistics/mortgage-arrears/residential-mortgage-arrears-and-repossessions-statistics-december-2018.pdf?sfvrsn=4

banking profitability. The residential mortgage portfolio made up approximately one third of the €30Bn of adverse scenario expected losses predicted at Irish banks between 2011 and 2013 (Prudential Capital Assessment Review).

The existence of NPE's also directly affect Irish banking profitability through provisioning. In an article published as part of the Central Bank of Ireland's Quarterly Bulletin report in 2018, Donnery et.al (2018)[5] discuss how the level of provisions that are tied up in a non-interest earning NPE equates to money that is not earning interest on a performing loan an claim that the aim for central banks is for financial institutions to generate sustainable profits, which is negated by the existence of NPEs. However, a reduction in NPEs, will facilitate improved profitability, which in turn leads growth in capital, putting financial institutions in a stronger position to meet regulatory requirement.[6]

2.2.4 Factors Influencing Mortgage Arrears

The literature in this area broadly focuses on mortgage default as distinct from mortgage arrears. Given that mortgage arrears are precursor to mortgage default the factors studied in the literature for defaults are considered herein applicable to mortgage arrears in the context of this study.

Gerlach-Kristena & Lyons (2017) examined the drivers of mortgage arrears across Europe over the period 2004-2011, or more specifically the factors impacting the inability of mortgage repayment. Notwithstanding age, education and other household specific characteristics, for which the dataset was normalised and controlled, through regression analysis the authors find that disposable income and high mortgage payments, namely loan affordability, are the most significant factors impacting mortgage arrears and repayments for short term arrears.

Connor and Flavin's (2015) research related to an examination of the causes of mortgage default using a dataset of mortgage loans held with Permanent TSB in September 2013 relying on six explanatory variables for loan default. They concluded that unaffordability variates play a significant role in mortgage default. Kelly (2012) provided a framework

[5] http://cdn.thejournal.ie/media/2018/04/resolving-non-performing-loans-in-ireland-2010-2018.pdf

[6] https://www.centralbank.ie/news/article/transforming-banking-for-customers-a-regulatory-perspective---deputy-governor-ed-sibley

for estimating probabilities of default of individual mortgages in Ireland in a multi-state Markov model, finding that macro-economic factors such as the house price index and unemployment rates were significant factors.

These factors impact borrowers' ability to pay. “Ability to pay” is a widely accepted theory surrounding the decision of borrowers to fall into arrears or loan default. In general terms, borrowers will not fall into arrears provided their disposable, non-household related, income remains sufficient to meet loan repayments without undue financial burden. (Whitley, Windram, Cox, 2004) It is noted that “shocks” to the ability to repay result in non-payments and accumulations of large arrears balance (Kelly & McCann, 2016), as observed in the case of Ireland's arrears portfolio presented earlier.

This has significance for the Irish banking sector in lowering arrears levels and preventing defaults into the future. If banks had the ability to assess the use, and therefore availability of disposable income of PDH/BTL loan customers, loan affordability from the customer's viewpoint could also be monitored and kept on track. This allows banking institutions to proactively monitor customer issues for the timely identification of potential vulnerabilities, in the guise of Early Warning Indicators. In turn, this may be used as a risk-management tool to prevent the risk of such accounts falling into arrears through early intervention. The Irish Banking Federation has also noted the importance of early customer engagement in the management of arrears.

While the level of disposable income available to households has been identified as a driver in loan affordability, research is limited on the specific factors impacting loan affordability and how the end uses of disposable income may impact loan affordability into the future. This study therefore distinguishes itself from previous work both by focusing on the use of transactional level data as a predictor variable, and in the utilisation of real time data made available by lender A, which contains information on 50,436 distinct live accounts, rather than survey data or loan level data.

2.2.5 Policy & Guidance

A key responsibility of the Central Bank surrounding the approach to mortgage arrears resolution is geared towards safeguarding the fair treatment of customers. This is guaranteed through a strong consumer protection framework while ensuring banks also have appropriate arrears resolution strategies and operations in place.

In the aftermath of the financial crisis in Ireland, the Central Bank introduced several measures to mitigate the risk of such events recurring. These included:

- The introduction of borrower-based measures that limit loan-to-value and loan-to-income ratios, which increase the resilience in the system and reduce the risk credit-fuelled property bubbles from over-borrowing and over-lending;
- The Consumer Protection Code and the Code of Conduct on Mortgage Arrears (CCMA) which govern how lenders interact with retail borrowers that are in distress.
- The Mortgage Arrears Resolution Process (MARP) which must be followed when dealing with customers facing arrears.

The key message of the Financial Conduct Authority's 'Early arrears management in unsecured lending' Thematic Review is that firms need to do more culturally to identify vulnerable customers and deliver fair outcomes. This is true too of the banking sector and in addition to domestic measures, the European Central Bank (ECB) has defined guidelines[7] (2017) to reduce NPE levels across Europe.

According to the ECB, Ireland has been recognised as one of the poorest performing member states, and while there has been a decline in NPEs as evidenced in figure 2.1, the Irish banking sector is required to conform to ECB guidelines and adopt the advisory measures to further reduce and prevent NPEs. One such measure is the implementation and identification of Early Warning Indicators (EWIs). EWIs are a set of indicators that aim to detect potential credit deterioration before negative events occur. The guidance set out by the ECB states that banks should determine EWIs at a number of levels including portfolio and customer transaction level. Behavioural scoring systems are just one of the examples of transactional level EWIs provided in the guidance.

2.3 Credit Scoring

Credit Risk is defined as the risk of loss arising due to any real or perceived change in a customer’s ability or willingness to repay their financial obligation (Anderson, 2007). Financial institutions use a classification technique broadly termed as “credit scoring” to evaluate the credit risks related with lending to a customer. The purpose of credit

[7] https://www.bankingsupervision.europa.eu/ecb/pub/pdf/guidance_on_npl.en.pdf

scoring is to assign scores to the characteristics of customers and historical default as an indication of the risk level of the borrower with the aim to build a single aggregated risk indicator for a set of factors (Bolton, 2009).

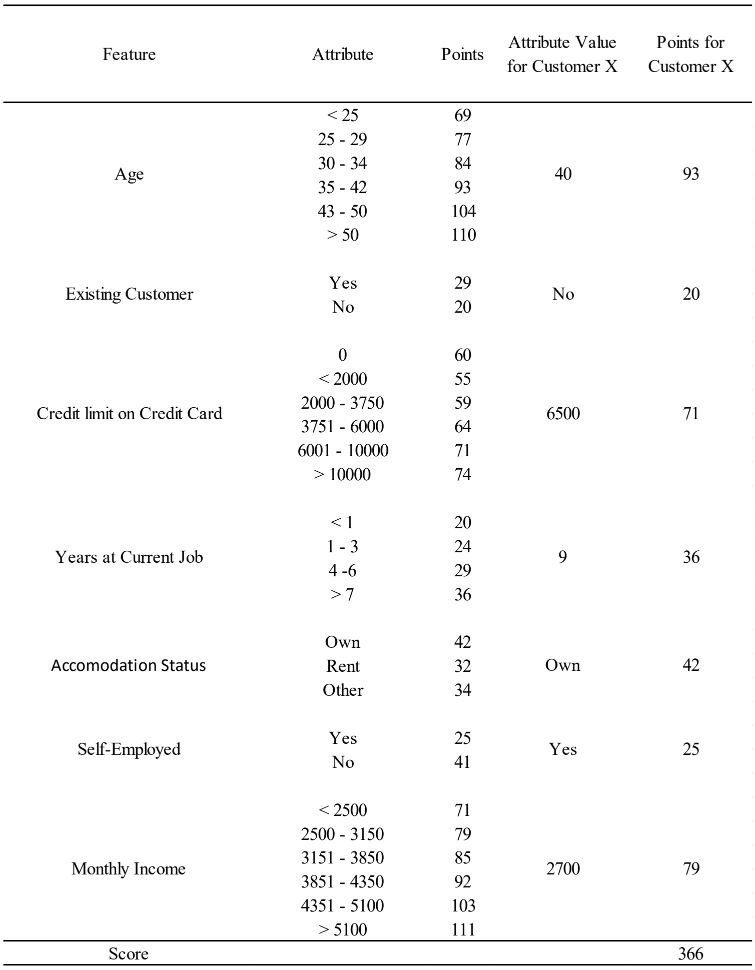
The key assumption made, which underpins the development of a credit scoring model, is that the future resembles the past. For historical customers, it is possible to analyse their past behaviours and distinguish them as one of two groups: good customers (those who will not default on their financial obligation) or bad customers (those that will default on their financial obligation). Credit scoring is essentially a method for classifying customers into these two groups.

Prior to the introduction of advanced credit scoring models, a qualitative, expert-based approach was taken to make a decision about credit risk. This was done by inspecting the five C’s of the customer (Baesens, Rosch & Scheule, 2019).

- Character – the customer’s reputation
- Capital – the difference between the customer’s assets and liabilities
- Collateral – the security being offered
- Capacity – the customer’s ability to pay
- Condition – the performance of the current economy

This qualitative approach had several short comings. It was unreliable, judgemental, not replicable and time consuming to reproduce for a large number of customers. With the emergence of statistical classification techniques in 1980s, financial institutions started utilising statistical approaches (Hand, 2001). Two key statistical approaches for credit scoring are application scoring and behavioural scoring. The purpose of application scoring is to predict, at the time of the loan application, the customer’s probability of defaulting at some time in the future. Variables including total liabilities, total debt, age, gender, marital status and income are generally used to build application scoring models. Behavioural scoring analyses the behaviour of existing customers who have secured credit from the lender. Because application scoring is initiated for customers with new loan applications and the type of features included are general demographic features, it will not form part of this research. Conversely, because behavioural scoring focuses on existing customers, it can be used to assess the capabilities of transactional data to monitor credit risk. Behavioural scoring is discussed in more detail in Section 2.3.1.

As mentioned previously, the aim of credit scoring is to build a single aggregated risk indicator for a set of factors. The most basic credit scorecard consists of a set of risk indicators that are statistically proven to be strong predictors of credit risk. An example of a scorecard is provided in figure 2.2 using features such as age, previous financial history, employment, credit card details and monthly income to assign the customer a credit score.



Feature	Attribute	Points	Attribute Value for Customer X	Points for Customer X
Age	< 25	69	40	93
	25 - 29	77		
	30 - 34	84		
	35 - 42	93		
	43 - 50	104		
	> 50	110		
Existing Customer	Yes	29	No	20
	No	20		
Credit limit on Credit Card	0	60	6500	71
	< 2000	55		
	2000 - 3750	59		
	3751 - 6000	64		
	6001 - 10000	71		
	> 10000	74		
Years at Current Job	< 1	20	9	36
	1 - 3	24		
	4 -6	29		
	> 7	36		
Accomodation Status	Own	42	Own	42
	Rent	32		
	Other	34		
Self-Employed	Yes	25	Yes	25
	No	41		
Monthly Income	< 2500	71	2700	79
	2500 - 3150	79		
	3151 - 3850	85		
	3851 - 4350	92		
	4351 - 5100	103		
	> 5100	111		
Score				366

Figure 2-2: Example Scorecard for Customer X

In figure 2.2, it can be seen that each feature is split into two or more attributes with a score generated for each attribute. The attributes with the higher scores are associated with customers who are, statistically, less likely to default. Customer X is assigned a score for each feature which are added together for an overall credit score.

2.3.1 Behavioural Scoring

In behavioural scoring, a population of customers is chosen so that the data on their performance is available either side of a pre-determined single observation point. The period before the observation point is called the observation window. The features that are recorded during the observation window are used to describe the customer's performance. The most common features used in behavioural scoring include average, maximum, minimum levels of balance, credit turnover, and debit turnover as well as indicators of delinquent behaviour e.g. number of missed payments, number of months where overdraft was exceeded (Thomas, Ho & Scherer, 2001).

The period after the observation point is called the outcome window. During this window, the customer is classified as being good or bad, depending on their default status. Anderson (2007), outlines two approaches which financial institutions can chose between when classifying their customers as good or bad: (i) *current status* approach which classifies customers based on their status at the end of the pre-defined outcome window or (ii) *worst status* approach which classifies a customer based on their worst status during the outcome window.

To build a behavioural scoring model, financial institutions are required to make decisions on a number of parameters outlined above. Firstly, the range of historical data from which to model customer performance i.e. the length of the observation window. Secondly, how far into the future does the financial institution want to make a prediction i.e. the length of the outcome window and finally a decision is required as to what defines a defaulter/bad customer. There is no standard approach to determining the length of the observation and outcome window. A review of literature shows recommendations which range from 6 to 24-month windows (Thomas et al., 2001; Thomas, 2009; van Gestel and Baesens, 2009). Kennedy et al (2013) evaluated the contrasting effects of altering the observation window and outcome window as well as consider the two approaches outlined above for classifying customers into good or bad categories. The results of their work indicated that a 12-month observation window and a 6-month outcome window yielded the best results. Additionally, they concluded that the worst status approach for classifying customers gives a higher assurance that the classification will be correct compared to the current status approach.

2.4 Transactional Data and Credit Risk Models

While there is a substantial amount of literature focused on the development and improvement of credit risk models, it is clear that there is a shortage of research into how transactional factors affect these models. Traditional features used in credit risk models include: (i) loan characteristics such as the loan amount, loan term, payment frequency; (ii) customer characteristics such as geographic location, age, number of children, monthly income and (iii) behavioural characteristics such as loan instalment amount, arrears trends, number of unpaid transactions (Galindo and Tamayo, 1997; Hand and Henley, 1997; Feldman and Gross, 2003; Kennedy et al. 2013; Kelly and O'Malley, 2015; Fitzpatrick and Mues, 2016)

However, Khandani, Adler and Lo (2010), in their paper, discuss that while the use of traditional features produces reasonably accurate results, these features "adjust slowly over time and are relatively insensitive to changes in market conditions". They argue that one of the most important drivers of macro-economic conditions and credit risk is customer spending. According to the Central Bank of Ireland's Credit and Debit Card Statistics, Point of Sale (PoS) credit and debit card expenditure (excluding ATM transactions) rose to a record high of €5.1 billion in December 2018[1] (Fig 2.3).

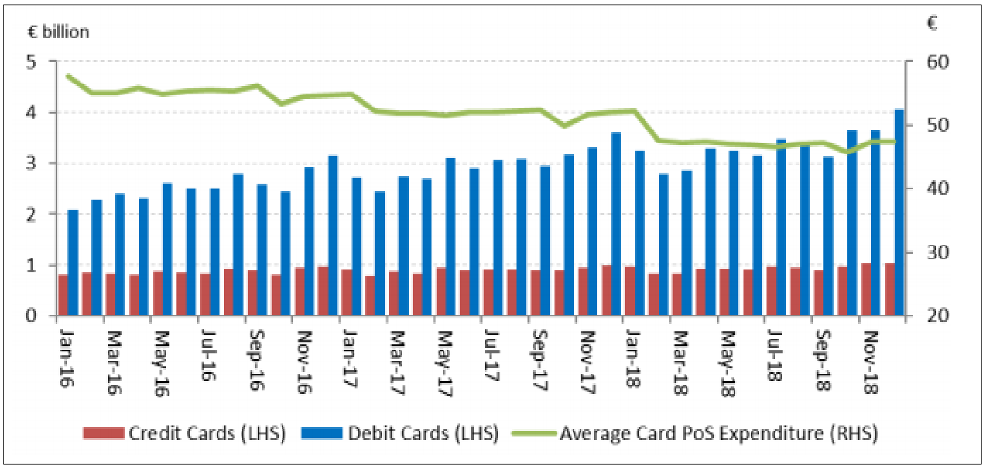


Figure 2-3: PoS Card Expenditure Ireland

Khandani, Adler and Lo (2010) conclude that, by including transactional features in a credit risk model along with the aforementioned traditional features, their results are indicative of considerably more powerful models of customer credit risk. This is in line with the literature discussed in section 2.2.4 which points to disposable income being significant. Namely, the use of disposable income, and therefore the availability of it over time, as a transactional feature is likely to improve traditional credit scoring.

2.5 Knowledge Discovery, Data Mining and Predictive Modelling

Knowledge Discovery can be defined as the "the nontrivial extraction of implicit previously unknown, and potentially useful information from data" (Frawley, Piatetsky-Shapiro and Matheus, 1992). Fayyad, Piatetsky-Shapiro and Smyth (1992) outline an approach called the Knowledge Discovery in Databases (KDD) process. This is an iterative process centred on identifying patterns from data.

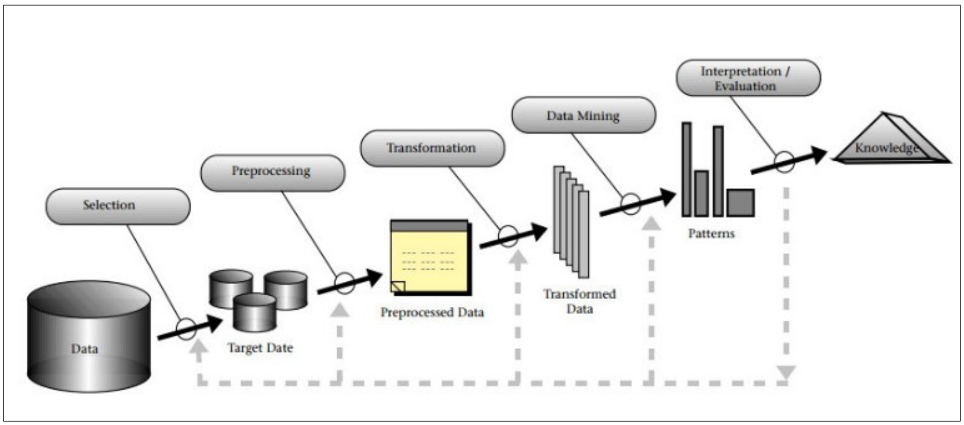


Figure 2-4: KDD Process (Source: Fayyad, Piatetsky-Shapiro and Smyth, 1992)

Such is the importance of data mining, it too has been subject to the development of methodologies and frameworks. The most widely used in the field of data analytics is the Cross Industry Standard Process for Data Mining or CRISP-DM for short (Shearer, 2000). Figure 2.5 illustrates the CRISP-DM process and how it is divided into six main steps: business understanding, data understanding, data preparation, modelling, evaluation and deployment.

Other frameworks for data mining have been developed including the Sample, Explore, Modify, Model and Assess or SEMMA (Azevedo & Santos, 2008). This framework is seen more as sequential steps that can be used for data mining. Therein lies the advantages of the CRISP-DM method. As illustrated in Figure 2.5, there is no restriction on moving between the different steps with the arrow wrapping around the whole process suggesting an iterative process that does not end at Deployment.

This research is centred on the predictive modelling component of the CRISP-DM process. Predictive modelling is the process of using historical data to predict future events. Predictive models are built by using a set of features which can be in any form e.g. numerical and/or categorical to predict a target or dependent variable. Like the predictive features, the target variable can either be numerical or categorical. If the target variable is numerical (or continuous), a predictive modelling technique called Regression is used and if the target variable is categorical, a technique called Classification is used. The models are trained using historic real-world data and tested using separate "unseen" data to evaluate their performance.

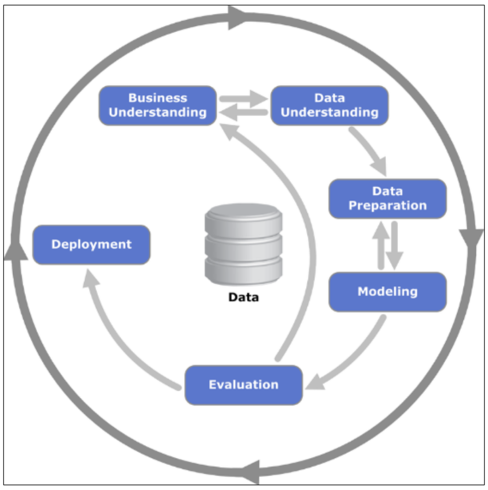


Figure 2-5: CRISP-DM Process Model (Source: Shearer, 2000)

2.6 Predictive Models for Classification

In this section, logistic regression, a typical classification algorithm used for building binary classification models, is discussed. Competing algorithms include k-nearest neighbours (K-NN), decision trees and Support Vector Machines (SVM). While this is not a complete list of classification techniques, it does include all that are suitable for use in financial institutions due to various regulatory compliance. This section will conclude with a review of the literature supporting the decision to use logistic regression.

2.6.1 Regression

Linear regression is a statistical technique that can be used to analyse the relationship between a single dependent (or target) variable and one or more independent (or predictor) variables.

In its simplest form, linear regression aims to predict the dependent variable, using a single independent variable. This is called simple linear regression and is often represented by the following equation:

$$Y = b_0 + b_1 X$$

where Y is the dependent variable, X is the independent variable, b_0 is the intercept (the value of Y when X = 0) and b_1 is the slope of the line.

A straightforward example of simple linear regression is illustrated in Figure 2.6. It shows the relationship between the number of credit cards a family holds and the family size. In this example, the dependent variable, Y, is the number of credit cards a family holds and the independent variable, X, is the family size.

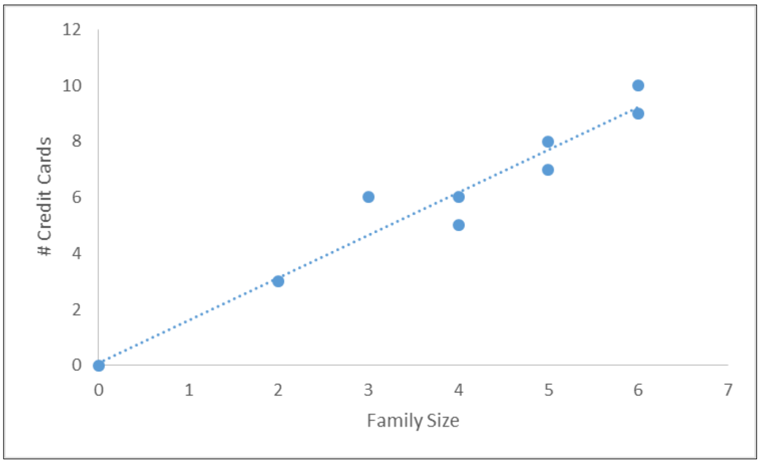


Figure 2-6: Simple Linear Regression

Examples such as the above can be expanded on to include additional independent variables. This is known as multiple linear regression and can be illustrated using the below equation:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n$$

As with linear regression, the goal of logistic regression is to find the best fitting model to describe the relationship between a dependent variable and one or more independent variables (Hosmer and Lemeshow, 2013). What distinguishes logistic regression from linear regression is the dependent variable. Linear regression models are sufficient to use when the dependent variable is continuous $[-\infty, +\infty]$. Because the dependent variable in this research is binary, i.e. there are two possible outcomes (1 for default, 0 for non-default), it is necessary to use logistic regression. One of the advantages of logistic regression is that it does not require the data to be normally distributed and there is no requirement for the dependent variable and independent variables to be have a linear relationship (Hosmer and Lemeshow, 2013).

The logistic function describes the mathematical form on which logistic regression models are based (Kleinbaum, Klein & Pryor, 2002). The logistic function is represented by the following equation:

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

Figure 2.6 illustrates how this function is plotted as z varies from $-\infty$ to $+\infty$. Note, that as z gets closer to $-\infty$, the logistic function equals 0 and as z gets closer to $+\infty$, the logistic function equals 1. In other words, the range of the logistic function will always be between 0 and 1, regardless of the value of z as represented by the S-shape curve plotted in figure 2.7. It is for this reason logistic regression models are so popular (Kleinbaum, Klein & Pryor, 2002). Logistic regression is designed to describe a probability, which is always in the range of 0 and 1. For this research, such a probability provides the risk of a customer going into default.

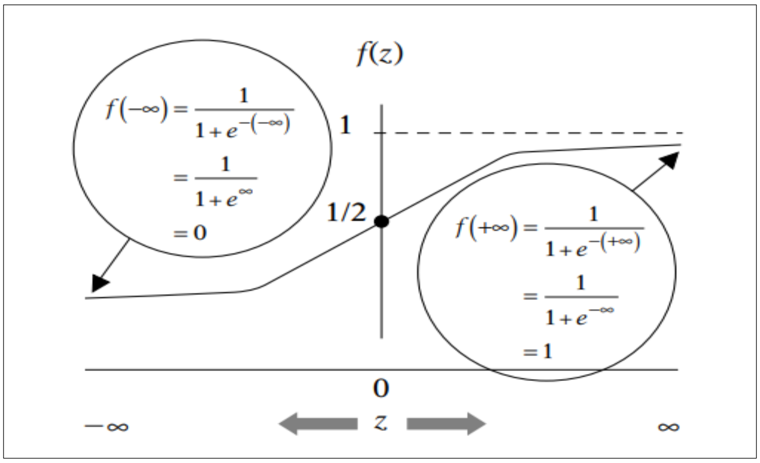

f(z)
$f(-\infty) = \frac{1}{1+e^{-(-\infty)}}$
$= \frac{1}{1+e^{\infty}}$
$= 0$
1
1/2
$f(+\infty) = \frac{1}{1+e^{-(+\infty)}}$
$= \frac{1}{1+e^{-\infty}}$
$= 1$
0
$-\infty$
z
∞

Figure 2-7: Logistic Function (Source: Kleinbaum, Klein & Pryor, 2002)

In order to obtain the logistic regression model from the logistic function, z is substituted with the linear regression equation.

$$f(z) = \frac{1}{1 + e^{-(b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n)}}$$

2.6.2 Industry Standard

Logistic regression is the most used technique for credit scoring (Bolton, 2009). While there is extensive research on a number of additional classification techniques including, amongst others, decision trees and K-NNs (Henley & Hand, 1997; Galindo & Tamayo, 1997; Brown & Mues, 2012), these techniques have not been widely used in developing credit scorecards in financial institutions (Thomas, 2009). There are two main reasons for this. Firstly, these techniques lack robustness. Methods like neural networks and support vector machines are more vulnerable as the characteristics of the population change (Dong, Keung Lai & Yen, 2010). Secondly, there is a lack of transparency with the aforementioned techniques. Regulators require financial institutions to provide any reason for rejecting a customer for credit (Thomas, 2009). Because these techniques do not require information about the relationships between variables, their results are difficult to interpret and hence, financial institutions are unable to provide reasons for rejecting credit based on these results.

Logistic regression is not impacted by either of the above issues. However, its prediction power is slightly inferior to some of the other classification types (Galindo & Tamayo, 1997; Brown & Mues, 2012). Therefore, in this paper, the addition of transactional features along with the examination of observation and outcome window lengths is proposed to improve the prediction accuracy of logistic regression.

2.7 Variable Selection Methods

According to Guyon and Elisseeff (2003), there are three objectives of variable selection; (i) improving the prediction performance of the predictor variables, (ii) providing faster and more cost-effective predictors and (iii) providing a better understanding of the underlying process that generates the data.

In their paper, Guyon and Elisseeff (2003) divide the approaches for selecting variables into three methods; wrapper method, filter method and embedded method.

The wrapper methodology uses a subset of variables and trains the model using these. It then iterates through that process constantly adding and removing variables returning the optimum result and variables used (Kohavi and Sommerfield, 1995). Koller and Sahami (1996) examined methods for variable subset selection and part of their conclusions highlighted the extremely high computational cost of using the wrapper method for variable selection. This has been highlighted as an issue throughout the literature (Chandrashekar & Sahin, 2013; Kumar & Minz, 2014).

The filter approach selects a variable subset as a pre-processing step. The approach used variable ranking techniques such as correlation coefficients, information gain and distance measures, amongst others (Chandrashekar & Sahin, 2013). Research shows that filter methods are more practical than wrappers because they are quicker to compute (Hall, 2000; Sánchez-Marono, Alonso-Betanzos & Tombilla-Sanromán, 2007)

Finally, the embedded approach combines both the wrapper and filter methods with the aim if reducing the computational time taken of the wrapper method. This is commonly achieved by including the variable selection as part of the training process (Guyon and Elisseeff, 2003).

In credit risk models, variable selection is important due to the large number of variables present in a credit scoring dataset (typically more than 100 variables) and the need to identify an effective subset of 10-20 variables (Hand & Henley, 1997). Anderson (2007) states that when deciding variables for inclusion in a scorecard, the main variables to consider should be logical, have a significant degree of predictive power, have a low correlation with each other and result in unacceptable information loss if excluded. Anderson (2007) proposes a mix of filter and wrapper methods for variable selection in credit scoring models. Hand and Henley (1997) support this proposition as well as taking into consideration domain or expert knowledge.

2.8 Class Imbalance

Class imbalance occurs when the number of records for each classification of the target variable is uneven i.e. when one class is represented by a large number of examples (the majority class) while the other class is represented only by a few examples (the minority class) (Japkowicz, 2000). In credit scoring, this is known as low default portfolios (LDP) where there are a much smaller number of observations in the default class than in the non-default class.

A number of techniques have been evaluated in the literature in order to identify an effective method of overcoming class imbalance. These include random under-sampling of the majority class, random over-sampling of the minority class and synthetic sampling of the minority class (Chawla, Bowyer, Hall & Kegelmeyer, 2002). Random under-sampling of the majority involves randomly sampling a portion of the population such that the number of observations in each class is more balanced. Conversely, random over-sampling and synthetic sampling of the minority involves duplicating or creating synthetic or new data based on the minority class to balance the observations in each target class. The two former methods have some shortcomings. For example, random under-sampling could potentially remove important samples while random over-sampling can lead to overfitting (Chawla, 2010).

There is extensive research on over and under-sampling methods to overcome the class imbalance problem (Chawla, Bowyer, Hall & Kegelmeyer, 2002; Japkowicz, 2000; Kubat & Matwin, 1997; Ling and Li, 1998). These studies have used different variations of over-sampling and under-sampling with sometimes conflicting views of the usefulness of under-sampling versus over-sampling.

Japkowicz (2000) discussed the effect of class imbalance by considering two sampling methods for both and under and over-sampling. For over-sampling, two resampling methods were considered. Random resampling entailed over sampling the minority class at random until it comprised of as many samples as the majority while "focused resampling" entailed over sampling only those minority samples that occurred on the boundary of minority and majority classes. Likewise, for under-sampling two down-sizing methods were considered. Random downsizing entailed eliminating random samples of the majority class until it matched the size of the minority class while "focused downsizing" meant removing only those samples that were furthest away. The author concluded that all methods used for over-sampling and under-sampling were equally effective methods for overcoming the class imbalance problem.

Ling and Li (1998) also combined under-sampling of the majority class and over-sampling of the minority class. They conducted three experiments. Firstly, they under-sampled the majority class and concluded that the greatest accuracy was obtained when the majority and minority classes were equally balanced. Secondly, they over-sampled the minority class with replacement to even out the number of majority and minority samples. They concluded that this combination did not have a significant effect on increasing the accuracy.

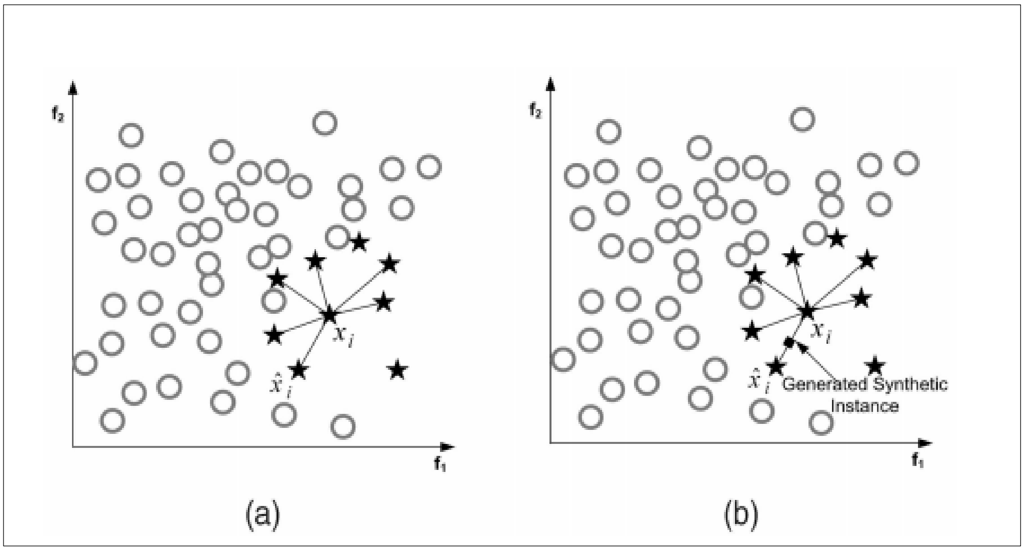


Figure 2-8: SMOTE procedure (Source: He & Garcia, 2009)

Chawla, Bowyer, Hall and Kegelmeyer (2002) introduced a synthetic minority over-sampling technique (SMOTE). This approach also involved a combination of over-sampling the minority class and under-sampling the majority class. The method used to over sample the minority class involved creating synthetic minority class samples along the line segments joining a subset or all of the k minority classes. Figure 2.8 illustrates the SMOTE procedure where the stars and circles represent samples of the minority and majority classes respectively. He and Garcia (2009) highlight that SMOTE expand the dataset in a way that generally improves the model but also highlight some drawbacks of the technique including over generalisation and variance. Because of the shortcomings of both over and under sampling and also due to the inconclusiveness of literature in the field, there is merit in trialling both sampling techniques in this research to identify the most appropriate methodology.

2.9 Validation and Evaluation

2.9.1 Validation

An important step in developing a predictive model is evaluating the model. There are some important considerations to make prior to building a model in order for it to achieve optimal performance on unseen data. Referred to as *Resubstitution Validation* (Doughtery, Jianping & Bittner, 2007), building a model and testing the accuracy of the model using the same historical dataset may introduce an element of bias as the model may be over-fitted to the training dataset.

There are a number of methods that can be utilised in order to reduce this bias. One such method is the *holdout* method. This method involves partitioning the data into two mutually exclusive datasets and using one set to train the data (known as the training set) and the second set to test the data (known as the test set). It is common to use a 70/30 split in favour of the training set (Kohavi, 1995). Fig 2.9 illustrates how the holdout method can also incorporate a third subset (known as the validation set). While the validation set is not always required, it is useful for fine tuning the model.

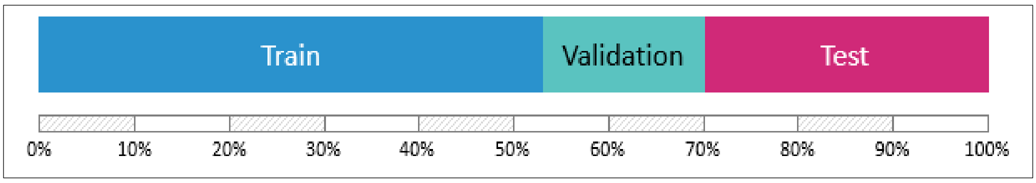
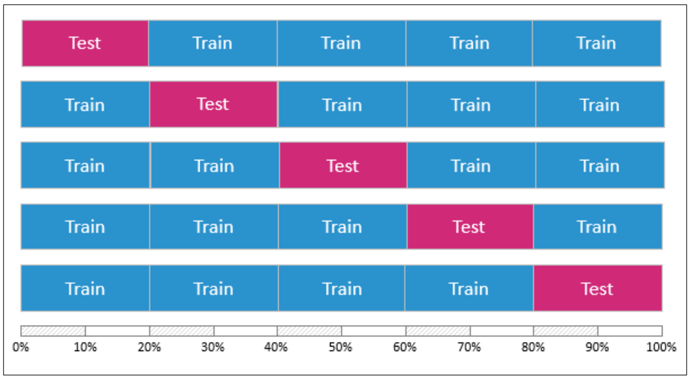


Figure 2-9: Holdout Validation Method

Another method used in validation is *K-fold cross validation*. Here, the dataset is split into k equal sized distinct subsets and iteratively trained on k-1 of these subsets and tested on the one remaining subset that has been excluded from training. The model is trained k times such that all samples are utilised as training and test sets throughout model building. An example of this can be seen in Fig 2.10. In his paper, Kohavi (1995) compared a number of validation methods including holdout and k-fold cross validation. The author concluded that a stratified approach to k-fold cross validation with $k = 10$ produced the best results.



Test	Train	Train	Train	Train
Train	Test	Train	Train	Train
Train	Train	Test	Train	Train
Train	Train	Train	Test	Train
Train	Train	Train	Train	Test

Figure 2-10: K-fold Cross Validation

2.9.2 Evaluation

The results of classification models map the modelled data into a specific category. Prior to this happening, a score cut-off or threshold must be set so that, as in this research (a binary classifier), a case with a score below the cut-off (closer to 0) is classified as non-default and a case with a score above the cut-off (closer to 1) is classified as default. The results of a binary classification model, such as the one in this research, can be observed by computing the number of cases correctly classified as non-default (true positives), the number of cases correctly classified as default (true negatives), the number of default cases incorrectly classified as non-default (false positive), and the number of non-default cases incorrectly classified as default (false negative). These four counts are generally illustrated using a confusion matrix, illustrated in table 2.1 (Sokolova & Lapalme, 2009)

		Predicted Class	
		Positive	**Negative**
Target Class	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Table 2-1: Confusion Matrix

Frequently used performance measures which can be calculated using the counts available in a confusion matrix include accuracy, misclassification rate, recall and precision. Details on how these are calculated can be seen in Table 2.3.

Of these, accuracy and the misclassification rate, or error rate, are the most popular measures used for assessing classification model performance (Hand, 2001). However, depending on the distribution of the target variable, both of these can be misguiding measures for model performance. Taking this research as an example, where the objective is to predict customers who will default with possible outcome values of “N” or “Y”. If 90% of the population were classed as “N”, it would be possible to create a model that has an accuracy of 90% by simply predicting class “N” for every sample as we can see in table 2.2. It is evident, therefore, that accuracy does not help with the research problem as the model has unsuccessfully predicted any customers who would default.

		Predicted Class	
		Positive	**Negative**
Target Class	**Positive**	0	0
	Negative	0.1	0.9

Table 2-2: Example Confusion Matrix

To alleviate this problem, there is a need for additional measures of model performance to be examined. Those measures include Recall, Precision and Average Class Accuracy. Recall (or Sensitivity) is a performance measure which evaluates the effectiveness of the model to identify positive cases (class "Y" in our example above). Precision tells us how often the model was correct when predicting a positive case. In the confusion matrix shown in table 2.2, both of these measures would have scored 0, highlighting the importance of these measures in particular examples such as the one in this research. Average class accuracy utilises both recall and specificity (the effectiveness of the model to predict negative cases), thus removing the effect of class imbalance (Bordersen et.al, 2010).

Measure	Formula	Description
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Measures the number of predictions that are correct
Misclassification Rate	$1 - Accuracy$	Measures the number of predictions that were incorrectly classified
Recall	$\frac{TP}{TP + FN}$	Effectiveness of classifier to identify positive cases
Precision	$\frac{TP}{TP + FP}$	Ratio of correctly predicted positive cases to the total predicted positive cases
Specificity	$\frac{TN}{TN + FP}$	Effectiveness of classifier to identify negative cases
Average Class Accuracy	$\frac{Recall + Specificity}{2}$	Average accuracy of positive and negative classes

Table 2-3: Measures for binary classification using notation from Table 2.1

Another useful method for evaluating how well the trained model fits the test data is the receiver operating characteristic (ROC) chart. A ROC chart plots the true positive rate (Recall) and the false positive rate (1-Specificity) over a number of threshold values. To compare ROC chart results of multiple classifier models, the area under the ROC curve (AUC) is used (Bradley, 1997). In the case of perfect classification, the AUC would take a value of 1 and for a random classifier (i.e. where the decision could ultimately be decided by flicking a coin), the AUC would be 0.5.

2.9.3 Statistical Tests

In his paper, Dietterich (1997) examined the use of statistical hypothesis tests to compare classifiers. The author recommended the use of the McNemar test, particularly in those cases where the algorithms that are being compared can only be evaluated once i.e. on one test set.

The McNemar test is a paired nonparametric test which operates on a contingency table by checking if the disagreements between the two algorithms are statistically different. The McNemar test statistic is calculated as follows:

$$\frac{\left((C1_y + C2_n) - (C1_n + C2_y)\right)^2}{(C1_y + C2_n) + (C1_n + C2_y)}$$

where:

- $C1$ refers to classifier 1
- $C2$ refers to classifier 2
- $C1_y$ refers to the number of instances that classifier 1 got correct
- $C1_n$ refers to the number of instances that classifier 1 got incorrect
- $C2_y$ refers to the number of instances that classifier 2 got correct
- $C2_n$ refers to the number of instances that classifier 2 got incorrect

2.10 Conclusions

This chapter presented a summary of the relevant literature surrounding mortgage arrears, the factors which influence them and binary classification in credit scoring, with a discussion of the use of transactional features in a credit risk model.

There has been a significant increase in the number of mortgage arrears in Ireland due to the 2008 banking crisis and even though there has been a steady decline in the level of arrears since 2014, they still continue to cause distress to borrowers and vulnerabilities to lenders. According to the research, the factors relating to mortgage arrears and ultimately default are several fold, with most authors agreeing to disposable income and high mortgage payments being key drivers. Additionally, guidelines set out by the European Central Bank instructed financial institutions to adopt measures to further reduce and prevent loans defaulting, including the implementation and identification of Early Warning Indicators (EWIs) and early intervention with customers facing potential arrears.

Financial institutions employ credit scorecards to evaluate the risk of existing or new customers defaulting on their financial obligation. There are generally two approaches to credit scoring, application and behavioural. Application credit scoring analysis typically utilises demographic information about the customer gathered from their loan application. Behavioural credit scoring analysis utilises information such as repayment behaviour and current account turnover.

The challenges financial institutions make when developing scorecards were discussed in detail. Challenges such as deciding what length the observation and outcome window should be and also deciding when to define a bad customer. The review of the literature highlighted no standard approach to these challenges with recommendations of 6 to 24-month windows (Thomas et al., 2001; Thomas, 2009; van Gestel and Baesens, 2009).

Throughout the literature, it is also acknowledged that there is a lack of research on the impact of transactional features in credit risk models. One study (Khandani, Adler & Lo, 2010) demonstrated that one of the most important drivers of credit risk is customer spending and produced considerable improved results in their credit risk model with the addition of transactional features. This study therefore aims to contribute to the literature by assessing the use of transactional features in predicting the default status of mortgage customers,

A large number of topics were discussed in the field of data mining and predictive modelling such as logistic regression, variable selection, the class imbalance problem, and evaluation and performance measures. Logistic regression is the most commonly used algorithm in the industry of credit risk modelling due to its transparency and robustness and will therefore be used to address the objectives for this research.

3 EXPERIMENT DESIGN AND METHODOLOGY

3.1 Introduction

This chapter presents the design and methodology of the experiments that were carried out as part of the research as well as present the data that was used throughout this research. There are two main sections in this chapter.

The first section provides the background on the set up of the experiments as well as outlines where the population of customer data was sourced from and what criteria was used during the selection process. This section also outlines what transactional features were created as part of the research, how these features were created and what pre-processing, transformation and data wrangling techniques were required in order for the features to map into a single analytical base table (ABT).

The second section focuses on the design and methodology undertaken for the modelling phase of the research. Techniques chosen for feature selection, sampling, validation and evaluation will be outlined in this section.

The chapter will conclude with a brief summary of the software used throughout the research.

3.2 Experiment Set-Up

The main objective of the research was to examine the predictive capability of transactional features in predicting whether mortgage customer will default or not. This was achieved by deriving and developing historical transactional features based on customer spending. Feature selection methods were applied to the transactional features to identify those which were highly predictive of customers defaulting on their mortgage repayment obligations. Prior to this, the issue of imbalanced datasets was highlighted. An experiment was undertaken on the baseline model which examined a number of sampling techniques.

3.3 Data

This section outlines the population of customers that were in scope for the experiments in the research. In Figure 3.1, mortgage customers were selected at six *observation points* from October 2017 to March 2018. These customers were not in default at the observation point. The window twelve months prior to each observation point is called the *observation window*. Data in this window was used for modelling purposes. Transactional data from this period was combined with existing credit risk model features from Lender A to form the trained data.

The twelve-month window directly after the observation point is called the *outcome window*. The customer is classified as being good or bad, depending on their default status during the outcome window. In section 2.3.1, two methods for establishing default customers during the outcome window are discussed: (i) the worst status approach and (ii) the current status approach. For the purpose of this research, the industry standard worst status approach was chosen, whereby if the customer is in default at any stage during the outcome window, they were classified as a default customer.

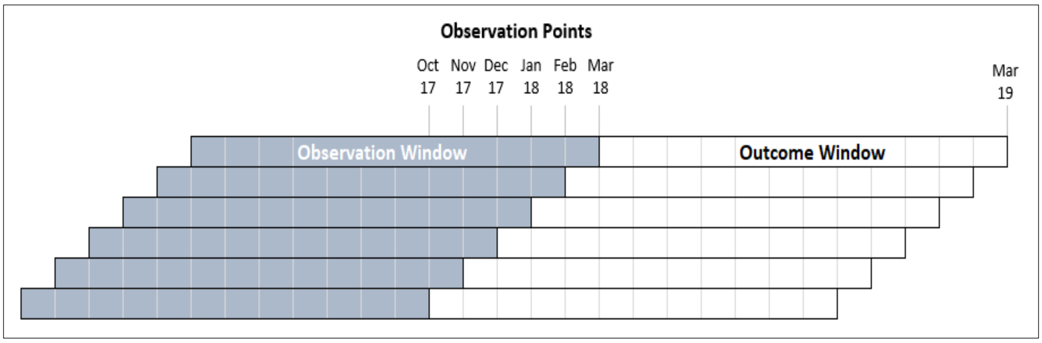


Figure 3-1: Observation and Outcome Window

The customer population used in this research was sourced from Lender A’s enterprise data warehouse (EDW). It contained details of 50,436 distinct customers who opened a mortgage account post 2013. By developing the population at a customer level, an element of duplication was introduced where, for example, a joint mortgage account existed with more than one customer associated with it. However, these customers had differing transactional histories on their associated current accounts. This was only a subset of mortgage customers on Lender A’s mortgage portfolio as only one of the loan systems within Lender A was selected for the research. The population of customers was observed at the six observation points highlighted in Figure 3.1, with data points, including eight existing scorecard features and derived transactional features, collated during each observation window. These features were combined to give one dataset to be used for training purposes. The current features used in internal credit risk models cannot be disclosed in the research paper due to sensitivity of the information. Table 3.1 outlines the make-up of the population at each observation point.

Observation Point	# Records	Non-Default	Default	Non-Default/Default Percentage
Oct 17	47,114	46,266	848	98:02
Nov 17	47,985	47,136	849	98:02
Dec 17	48,512	47,674	838	98:02
Jan 18	49,096	48,287	809	98:02
Feb 18	49,827	49,032	795	98:02
Mar 18	50,436	49,661	775	98:02
Total	**292,970**	**288,056**	**4,914**	**98:02**

Table 3-1: Characteristics of dataset for customer population

From Table 3.1, the class imbalance issue is evident with only 2% of the total population classified as being default by the end of the twelve-month outcome window. As discussed in Section 2.8, the class imbalance issue raises certain challenges when developing predictive models. Methods of dealing with this issue in the context of this study will be discussed in section 3.7.

3.4 Transactional Features

As discussed in the previous section, the transactional features were collated from each of the six observation windows. The transactional data was sourced from two separate databases in Lender A’s data warehouse: a VISA debit card transactional database and a non-card transactional database. In total approximately 270 million separate daily transactions were collected between all observation windows. Due to the large number of transactions, it was necessary to categorise these so that they were captured at a higher level.

In order to implement this categorisation of the transactional data, a merchant category code (MCC) was used. Merchant Category Codes are used external to Lender A to classify the primary business of a merchant. Within Lender A, these merchant category codes are used to create a Money Manager Application, allowing customers to keep track of their own spending behaviour. In total there are 284 merchant category codes which can be grouped up into four merchant category sectors each with a number of merchant category sub-sectors. A sample of these categories is illustrated in Table 3.2.

MCC	MCC Sector	MCC Sub-Sector	MCC Name
4119	Services	Health	Ambulance Services
4812	Services	Electrical Goods	Telecommunication Equipment
4814	Services	Utilities	Telecommunication Services
5013	Retail	Auto	Motor Vehicle Supplies
5047	Retail	Health	Medical Supplies
5411	Retail	Groceries	Grocery Stores, Supermarkets

Table 3-2: Merchant Category Codes

The availability of the merchant category codes within Lender A's data warehouse made it possible to categorise all transactions that took place on a Visa Debit card. For the non-card transactions, a set of rules were derived that placed the transactions into one of the merchant category codes and, where necessary, created a new category. These rules were developed in SQL based on a number of characteristics of the transactions including the transaction type (e.g. a transfer in or a withdrawal), the transaction source (e.g. ATM or branch) and the narrative associated with the transaction. To give an example, if the transaction type was flagged as a withdrawal and the transaction source was flagged as an ATM, this transaction was categorised as an ATM withdrawal. Additional information was gathered from keywords located in the narrative which were used to categorise the transactions, e.g. a narrative containing the word "bill" or "phone" was categorised as utilities. The final list of categories is listed in Table 3.3. The additional categories derived from the non-card transactions were ATM withdrawals, transfers in and out and savings.

Reference	Description
C101	Groceries
C102	Transfers In
C103	Withdrawals
C104	ATM Withdrawals
C105	Restaurants
C106	Auto
C107	Other Retail
C108	Utilities
C109	Clothing
C110	Insurance
C112	Savings
C113	Health
C114	Hardware
C116	Professional Services
C117	Entertainment
C119	Transport
C122	Accommodation
C129	Gambling
C132	Education

Table 3-3: Final list of Categories

In this experiment the categorised transactional data was used to identify trends in customer spending. For this reason, the transactional data was aggregated up to a monthly view, ending up with 12 months of a transactional spending pattern for each customer at each observation point. Two types of features were created using this monthly transactional data. Firstly, the rate of change between each transaction category was calculated using the formula below:

$$Rate\ of\ Change = \frac{Current\ Month\ Spend\ (Cat\ 1) - Previous\ Month\ Spend(Cat\ 1)}{Previous\ Month\ Spend\ (Cat\ 1)}\ x\ 100$$

Secondly, the monthly spend in each category was viewed as a percentage of the total spend in the month:

$$Category\ Percentage = \frac{Current\ Month\ Spend\ (Cat\ 1)}{Current\ Month\ Spend\ (Total)}\ x\ 100$$

In total there were 491 features generated for this experiment. The details of these can be found in Appendix A.

3.5 Creating Analytical Base Tables

The previous sections outlined the collation, pre-processing and transformation steps required to have the data in order for modelling purposes. As mentioned, this data was collected from a number of different sources. Fig. 3.2 outlines the process for generating two analytical base tables (ABT); one for the baseline model, containing features currently used in Lender A's internal credit risk models and one for experimentation combining transactional features with the original features.

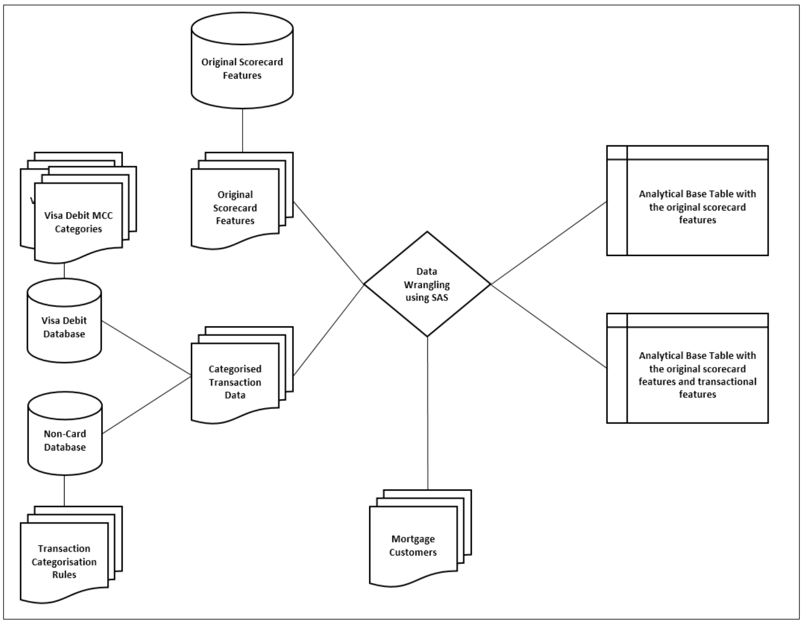

Original Scorecard Features
Visa Debit MCC Categories
Original Scorecard Features
Analytical Base Table with the original scorecard features
Data Wrangling using SAS
Visa Debit Database
Categorised Transaction Data
Analytical Base Table with the original scorecard features and transactional features
Non-Card Database
Transaction Categorisation Rules
Mortgage Customers

Figure 3-2: ABT development flowchart

3.6 Software

The experiment was designed and executed using SAS. SAS is a commercial tool commonly used in heavily regulated environments such as financial institutions and insurance companies. The reasons these companies use a commercial tool such as SAS as opposed to open source tools such as R and Python is because the analytical models developed are subject to external supervisory review and these would not meet requirements if developed in open source tools due to the lack of quality assurance and testing on the packages developed.

SAS offers a drag-and-drop interface, meaning users do not have to write any code, allowing for the building of models using the SEMMA methodology referenced in Section 2.5: sampling, exploration, modification, modelling and assessment. SAS also offers its own programming language which can also be used for data wrangling, sampling and model development. A useful component of SAS for this research was the ability to connect to Lender A's data warehouse. This was useful for the development of the ABTs through the use of defined macros and functions such as PROC SQL[8].

Additional tools and applications used included SQL for the development of the rules used to categorise the non-card transactional data and R to create visualisation. R is an open source statistical tool. It contains a number of libraries that can be used to create effective visualisations. The main library used in this research was "ggplot"[9].

3.7 Methodology

The following sections will outline the design and methodology for each experiment undertaken as part of this research. The section will begin by outlining the validation methodology used throughout the research as well as the evaluation techniques utilised. This will be followed by an outline of the steps taken to develop a baseline model which was used to compare with experimental models developed throughout the research. The first experiment undertaken focused on the class imbalance problem and evaluated a number of sampling methods with the aim of finding an optimal strategy to deal with the imbalanced data. The remaining sections will involve discussing methodologies for assessing the use of transactional features. Feature selection methods will be discussed and modelling strategies using these features will also be outlined. A methodology for validating and evaluating models is also discussed in this section.

[8] https://support.sas.com/resources/papers/proceedings/proceedings/sugi27/p191-27.pdf

[9] https://www.rdocumentation.org/packages/ggplot2/versions/3.1.1

3.7.1 Model Validation and Evaluation

The validation method used in this research was the *holdout* method as discussed in Section 2.9. This method involved splitting the data into a training and validation dataset and a test dataset. The training dataset was used to fit the model while the validation dataset was used to monitor and tune the model in order to improve its ability to adapt to unseen data.

Each model that was produced as part of this research was evaluated through the use of a number of performance measures. Model performance was evaluated on training, validation and test datasets. The expectation was that the models would perform well on the training dataset and perform in a similar fashion on the validation dataset. It would be considered a sign of overfitting if the model performance on the validation dataset performed considerably worse than the model performance on the training dataset.

There are various performance measures that can be used for classification problems. As discussed in Section 2.9, the most common measures used are accuracy and the misclassification rate. However, in the scenario of a classification problem where the data is imbalanced, both of these measures are not reliable and so alternative measures such as recall, precision and specificity must be considered. The rationale for choosing which measure to use relies on the cost associated with misclassifying records. In this experiment, there was a high cost associated with cases that were labelled as being in default but were being classified as not being in default (i.e. false negatives). Alternatively, the cost associated with cases that were labelled as not being in default and being classified as being in default (i.e. false positives) was much lower. For this reason, the performance measures chosen were recall and average class accuracy. Recall measures the effectiveness of the model to identify positive cases, in this case a positive case is relates to the default class. Average class accuracy measures the average accuracy of both the positive and negative cases. The AUC was also be used to compare models over a number of cut-off points.

3.7.2 Baseline Model

When building and evaluating predictive models, it is beneficial to have a baseline model so that model comparisons can be made. As one the main objectives of this research was to establish if transactional features improve on an existing credit risk model developed in Lender A, the baseline model for this research aimed to replicate this internal model with the intention to improve it.

As discussed in Section 3.3, the internal model incorporates eight existing features which cannot be disclosed due to sensitivity of information. These features were obtained and trained using the population associated with this research. The outputs of this model were compared against models created as part of the experiments and statistical tests were undertaken to evaluate differences between the models. The expectation was that by including transactional features in the existing model in Lender A, the performance of the model would improve.

For the purpose of building the baseline model, 50% of the population was randomly sampled, consisting of 146,485 instances. This sample was split into a training and validation set and a test set with 80% making up the training and validation element and 20% making up the test element. Details of the modelling data are illustrated in table 3.4.

	# Records	**Non-Default**	**Default**	**Non-Default/Default Percentage**
Train/Validation	117,189	115,223	1,966	98:02
Test	29,296	28,805	491	98:02
Total	**146,485**	**144,028**	**2,457**	**98:02**

Table 3-4: Training and Test Split

The 117,189 instances in the training and validation set were split using a ratio of 75:25 so that the training dataset made up 75% of the data and validation made up 25% of the data. In Lender A, random under sampling of the majority class is utilised to address the class imbalance problem. To align with this methodology, the majority class in the training and validation set was randomly under sampled, so that the number of non-default customers matched the number of default customers. The baseline model was developed and validated using the under sampled training and validation datasets and tested and evaluated using the test dataset created at the outset.

3.7.3 Class Imbalance

As illustrated in Table 3.1, the dataset was heavily imbalanced in favour of non-default customers with just 2% of the population flagged as being in default. Whilst this is a common and expected occurrence in a mortgage portfolio, it creates a problem for building predictive models. As discussed in section 3.7,1, Lender A alleviate this problem by randomly under sampling the majority class until there is an even number of default and non-default customers. This, however, may result in the loss of important data. To overcome this, a number of sampling techniques were trialled to evaluate an optimal strategy for dealing with class imbalance.

This experiment used the same population and data that was used for building the baseline model. By doing so, it allowed results from the test dataset to be used to compare models. The methodology for this experiment used random under sampling and random over sampling.

For random under sampling, the majority class in the training and validation datasets was under sampled so that there existed a ratio 95:05 in favour of the non-default customers. This involved taking all instances from the minority class and a sample of instances from the majority class. A model was developed and tested using the test set and performance measures were recorded for comparison purposes. This process was repeated, under-sampling a higher percentage of the majority class at each iteration such that the ratio between non-default and default became smaller. In total nine models were developed using random under sampling.

A similar methodology was used for random over sampling. For example, the minority class in the training and validation datasets was over sampled so that the ratio of non-default and default customers becomes 95:05. Here, all instances from the majority class were considered and an over sample of instances from the minority class were taken. Again, by increasing the over sampling percentage, this process was repeated such that the non-default and default ratio became smaller. Nine further models were developed using random over sampling.

All eighteen models were trained and validated using the training and validation dataset and tested using the same test dataset. Performance measures for each model were evaluated and compared against the baseline model in order to establish the optimal strategy to implement in future experiments.

The experiment process used for addressing the class imbalance issue is illustrated in figure 3.3.

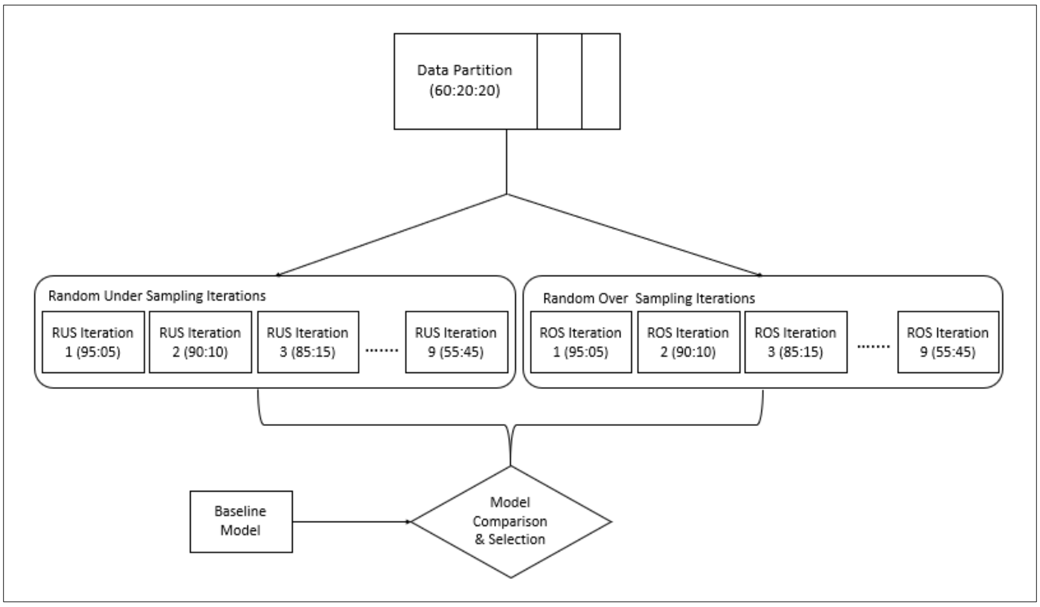


Figure 3-3: Process for Sampling Method Selection

3.7.4 Transactional Feature Selection

Following on from the results of the experiment outlined in section 3.7.3, this section will outline the methodologies of the experiment to evaluate the transactional features created as part of this research. Figure 3.4 illustrates this process at a high level.

This experiment incorporated the full population of customers and utilised the sampling technique which provided optimal results in the previous section. Stratified sampling was applied in order to build the training and validation dataset and test dataset. The training and validation dataset comprised of 80% of the data, while the test dataset comprised of the remaining 20% of the data. An updated baseline model was built on the full dataset, using the original credit risk features and the sampling technique chosen from the previous section. This was evaluated using recall, balanced accuracy and the AUC.

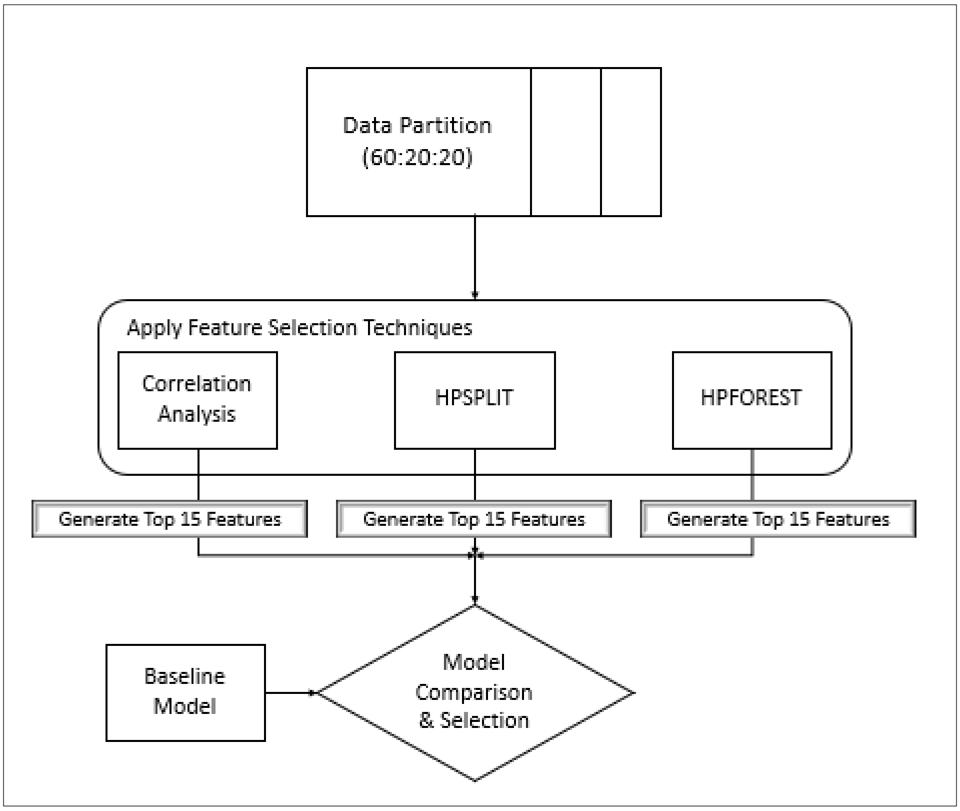


Figure 3-4: Process for Feature Selection

Once the baseline model was built, the next stage was to evaluate the derived transactional features. As part of the research, 491 transactional features were developed. Including such a high number of features in a model, can cause over-fitting and result in the model not generalising well on unseen data. As noted in section 2.7, a model should incorporate between 10 and 15 of the most effective features. Therefore, variable selection methods were used to determine which features had the strongest relationship with the target variable. Three methods of variable selection were explored; correlation feature based selection using PROC CORR[10] in SAS, decision tree variable selection using HPSPLIT[11] in SAS and random forest feature selection using HPFOREST[12] in SAS.

The correlation variable selection method was carried out using the training and validation dataset using the CORR procedure in SAS. Using the Pearson correlation coefficient, variables were assessed to ascertain how they correlated with each other.

[10] http://support.sas.com/documentation/cdl/en/procstat/66703/HTML/default/viewer.htm#procstat_corr_overview.htm

[11] https://support.sas.com/documentation/onlinedoc/stat/141/hpsplit.pdf

[12] https://support.sas.com/documentation/onlinedoc/hp-analytics-server/14/hpaug.pdf

Highly correlated variables were considered for removal with the variable having the strongest relationship with the target held for modelling purposes. Once the feature space was reduced, the top 15 features with the highest correlation with the target variable were included in the prediction model.

After the removal of highly correlated features, a decision tree variable selection method was examined using the HPSPLIT procedure in SAS. The HPSPLIT procedure builds decision tree models for both classification and regression. In this procedure, variable importance is calculated based on how each variable is used in the finished tree. Three metrics are used to establish variable importance; count, residual sum of squares (RSS), and relative importance. The count-based variable importance metric counts the number of times the variable is used in a split throughout the entire tree. The RSS based metric measures variable importance based on the change in the residual sum of squares. Finally, the relative importance metric is a number between 0 and 1 calculated by combining the RSS-based importance of a particular variable and the maximum RSS-based importance among all of the variable. For this research, relative importance was used as a variable selection technique. The top 15 features with the highest relative importance were recorded and included in the prediction model.

The final method used for variable selection was a random forest-based selection method using the HPFOREST procedure on SAS. This procedure calculates variable importance by evaluating the loss reduction for each variable. This procedure was used on the training and validation dataset after the removal of highly correlated features and again, the top 15 variables were selected for modelling.

After all of the three feature selection techniques were implemented, a set of models were developed, validated and tested incorporating the top 5, 10 and 15 variables from each method as well as the original factors used internally. A full evaluation was undertaken, comparing these models with the baseline model.

4 IMPLEMENTATION AND EVALUATION

4.1 Introduction

This chapter will discuss the implementation of the experiments being carried out as part of this research, to evaluate if transactional features are capable of improving existing credit risk models in Lender A. Exploratory data analysis will be undertaken on customer spending patterns and default trends.

The implementation of each experiment will be discussed sequentially, starting with the development of the baseline model based on the current credit risk model in Lender A. Results from each experiment will also be critically evaluated throughout.

4.2 Overview

The experiments carried out throughout this research aimed to evaluate the use of transactional features, derived based on customer spending, in predicting mortgage default. This was accomplished by retrieving a population of customers who have opened a mortgage account with Lender A and whose default status was known. The predictive model was built by incorporating the transactional features into an existing credit risk model developed in Lender A. Throughout the research, it was discussed that loan affordability is one of the key factors influencing mortgage arrears, which is driven by a customer's disposable income. With that in mind, the experiments in this study aimed to test whether non-traditional features, such as the transactional features, improved the existing models within Lender A which are presently based only on traditional features highlighted in section 2.3.1.

4.3 Data Exploration

This section will provide some exploratory data analysis and provide some summary statistics for the data. Due to the high volume of variables, only a portion of the data is described.

4.3.1 Mortgage Default

In total, there were 50,436 distinct customers with an open mortgage in Lender A considered for this analysis. Of those customers, 775 have a mortgage that has defaulted which equates to approximately 2% of the population. In Lender A, the overall default rate is higher. The difference here is the exclusion of customer whose mortgages opened pre 2013. These cohort of customers were excluded due to the considerable change in economy post-recession. In figure 4.1(a), the number of customers in default as a percentage of the total customers per county is represented. It is important to note that this map is not a fair representation of the country as a whole as it only takes into consideration the population studied as part of this research. Figure 4.1(b) shows the average household disposable income per person sourced from the Central Statistics Office[13].

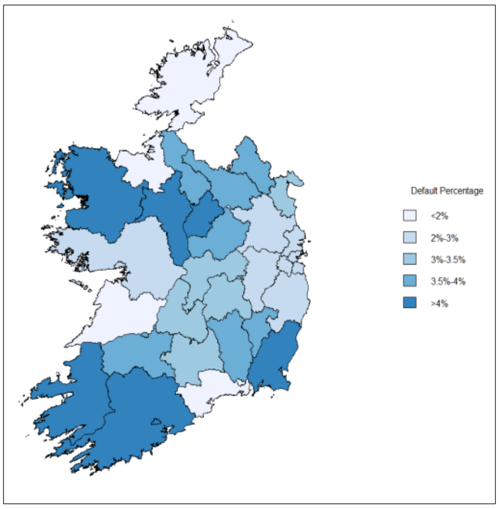


Figure 4-1(a): Percentage of mortgage customers in Default

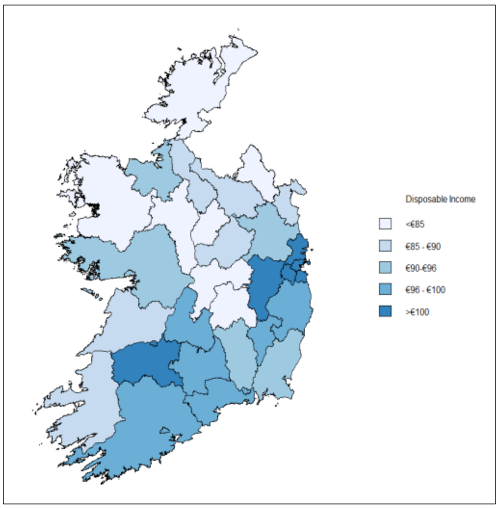


Figure 4-1(b): Household Disposable Income per person. Source(CSO)

For the most part, counties where a high average household disposable income per person is recorded generally have a low percentage of customers in default. Dublin, Kildare and Meath are good examples of this. Conversely, those customers with a low average household disposable income per person experiences a higher percentage of customers in default, e.g. Mayo, Roscommon, Cork and Kerry. This reconciles with the literature review where it was stated that disposable income and loan affordability are the key drivers of mortgage default.

[13] https://www.cso.ie/en/releasesandpublications/er/cirgdp/countyincomesandregionalgdp2016/

4.3.2 Average Spend per Category

The main objective of this research was to evaluate transactional features and their usefulness in terms of predicting mortgage default. As discussed in section 3.7.4, the transactional features were developed by creating and assigning transactions to a number of categories. This section will focus on a subsection of those categories with the purpose of highlighting the different spending patterns between those customers who did not default on their mortgage account obligations and those who did default on their mortgage account obligation.

Taking the full population, figure 4.2 (a), (b), (c) and (d) illustrate, for both non-default and default customers, the average monthly spend in four of the defined categories from section 3.4; insurance, ATM Withdrawals, gambling and saving. As expected, there were differences in the spending patterns for both sets of customers, with customers who defaulted spending, on average approximately €100 more than customers who did not default on a monthly basis.

The average spend on insurance for non-default customers was between €215 and €260 compared to that of the default customers who spent on average between €300 and €380. Similarly, default customers withdrew an average of €850 per month from ATMs compared to an average of €700 per month for non-default customers. Perhaps the most noticeable difference in monthly spending patterns were related to gambling and saving transactions. Default customer spending on gambling was almost twice that of non-default customer spending in most months. Likewise, non-default customers saved on an average €180 more than default customers over the twelve months. What was most noticeable was the behaviour of customers who subsequently defaulted with evidence of those customers withdrawing money from their savings account in months 6, 9 and 11. This indicates that default customers might not have had the capability to save or may have been experiencing cash flow issues in the months prior to defaulting.

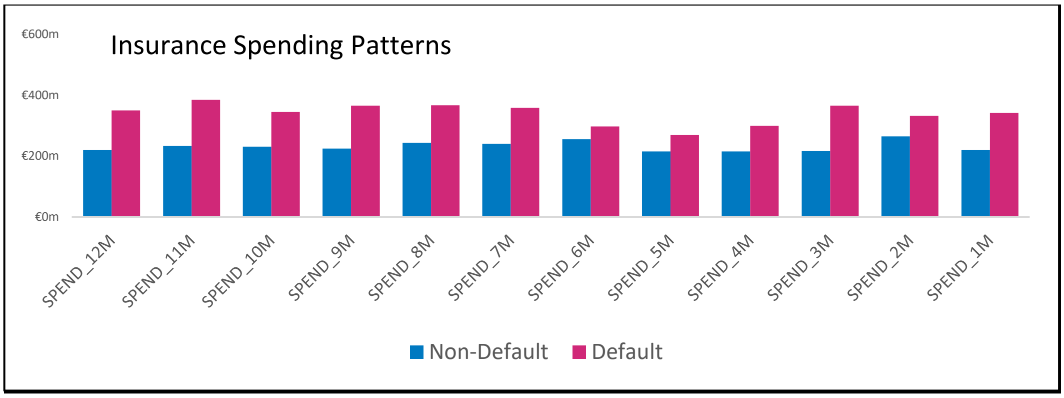


Figure 4-2(a): Average Monthly Insurance Spend by Default and Non-Default Customers

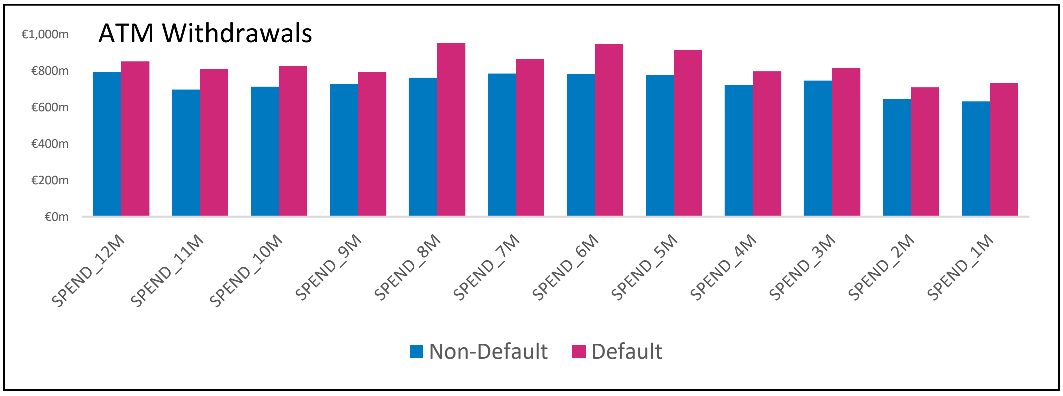


Figure 4-2(b): Average Monthly ATM Withdrawals by Default and Non-Default

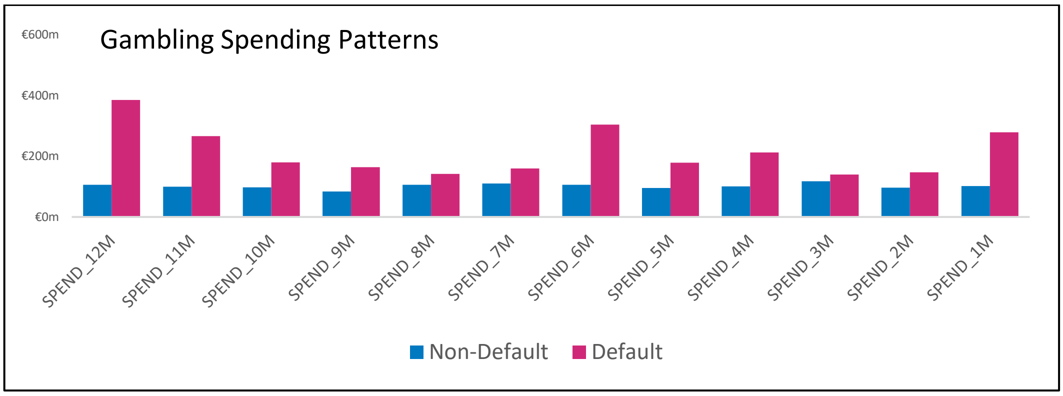


Figure 4-2(c): Average Monthly Gambling Spend by Default and Non-Default Customers

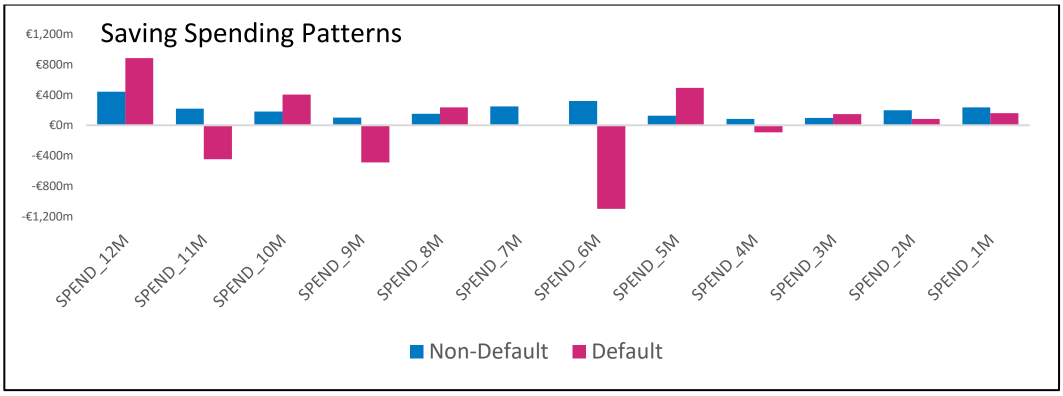


Figure 4-2(d): Average Monthly Saving by Default and Non-Default Customers

4.4 Baseline Model

As discussed in previous chapters, it is important to have a baseline model when developing predictive models. In this study, a baseline model was essential to compare and evaluate the results of experiments to follow. The baseline model for this research incorporated features from the existing credit risk model in Lender A. The following experiments firstly evaluated sampling techniques and secondly evaluated the use of transactional features. The results from these experiments were compared against the baseline model.

The model was trained in SAS using a random sample of 50% of the population, totalling 146,485 observations. Stratified sampling was used to split this sample into a training and validation set and a test set. Details of this split is illustrated in table 4.1.

	# Records	**Non-Default**	**Default**	**Non-Default/Default Percentage**
Train/Validation	117,189	115,223	1,966	98:02
Test	29,296	28,805	491	98:02
Total	**146,485**	**144,028**	**2,457**	**98:02**

Table 4-1: Sampling - Training and Test Split

Random under sampling was then applied to training and validation set to balance the percentage of default and non-default customers and it was divided into two separate datasets; 75% for training and 25% for validation. The baseline model was trained and validated using these two datasets. The model fitted to the training data with 75%

average class accuracy. When validated against the unseen test data, the model produced an average class accuracy of 68% with 59% Recall. The results would suggest that the model is not over-fitted. Additionally, the results found here are consistent with the models developed in Lender A. From here on in, this model will be referred to as $\mathrm{ML}_{\mathrm{BASE}}$.

4.5 Class Imbalance

There a number of methods which can be applied to the dataset to address the existing class imbalance. For this study, two methods were trialled; random under sampling and random over sampling. During the development of the baseline model, to keep it in line with the model development in Lender A, random under sampling was utilised so that the majority and minority class had an even 50:50 split. However, this may result in the loss of valuable information relating to customers who belong to the majority class i.e. not in default. The two methods trialled will be discussed below. For this experiment, the same dataset used to build the baseline model was used as outlined table 4.1. It is also important to note that while a model will benefit from training on a more evenly balanced dataset, it must be tested on a dataset which is a greater representation of the real world. Therefore, sampling methods were only applied to the training and validation dataset and tested using the test set. The expectation for this experiment was that a method of sampling where loss of data is minimised would perform better than the baseline.

4.5.1 Random Under Sampling

Random under sampling was implemented iteratively by randomly sampling the available customers whose mortgages were not in default, bringing the ratio of default to non-default customers closer to an equally balanced dataset at each iteration. For example, the first iteration sampled a percentage of the customers from the majority class such that the ratio of non-default to default customers became 95:05. Table 4.2 illustrates the characteristics of the training and validation dataset before and after each sampling iteration.

	# Records	Non-Default	Default	Non-Default/Default Percentage
Initial	117,189	115,223	1,966	98:02
Iteration 1	39,299	37,333	1,966	95:05
Iteration 2	19,711	17,745	1,966	90:10
Iteration 3	13,143	11,177	1,966	85:15
Iteration 4	9,802	7,836	1,966	80:20
Iteration 5	7,843	5,877	1,966	75:25
Iteration 6	6,575	4,609	1,966	70:30
Iteration 7	5,654	3,688	1,966	65:35
Iteration 8	4,962	2,996	1,966	60:40
Iteration 9	4,386	2,420	1,966	55:45

Table 4-2: Random Under Sampling

From table 4.2, it is clear the number of records from the majority, non-default class that are being lost at each iteration. Three models were trained and tested at each iteration using sampling without replacement. The average AUC, recall, and average class accuracy was recorded for each iteration. Table 4.3 details the results obtained for each of the three performance measures when the model was run using the test dataset where $ML_{RUS\_(x)}$ represents the results of the model which under samples the majority class such that the ratio of non-default to default is $100 - x{:}\,x$. For the most part there was considerable improvements across all performance measures at each iteration with AUC increasing from 0.7161 to a peak of 0.7405, recall increasing from 56% to a peak of 63% and average class accuracy increasing from 68% to a peak of 69%. Of the nine iterations, ML_{RUS_10} (under sampling majority such that the ratio of non-default to default is 90:10) outperformed the other sampling methods and the baseline model across all performance measures. ML_{Rus_10} fitted to the training data with an average class accuracy of 75% and 65% Recall. There was no evidence of over-fitting when compared to the results from the unseen test set highlighted in table 4.3. ML_{RUS_10} was compared with the random over sampling methods to determine which technique would be brought forward to be used in the next experiment. This is detailed in section 4.5.2.

Model Name	Model Description	AUC	Recall	Average Class Accuracy
ML_{base}	Baseline	0.7161	58.5%	67.8%
ML_{RUS_05}	95-05	0.7299	61.9%	69.1%
ML_{RUS_10}	90-10	0.7405	62.8%	69.4%
ML_{RUS_15}	85-15	0.7116	60.9%	68.7%
ML_{RUS_20}	80-20	0.7247	58.9%	68.7%
ML_{RUS_25}	75-25	0.7169	57.4%	67.6%
ML_{RUS_30}	70-30	0.7272	60.9%	68.8%
ML_{RUS_35}	65-35	0.7255	60.3%	69.0%
ML_{RUS_40}	60-40	0.7350	60.3%	69.1%
ML_{RUS_45}	55-45	0.7193	61.5%	68.4%

Table 4-3: Results from Random Under Sampling

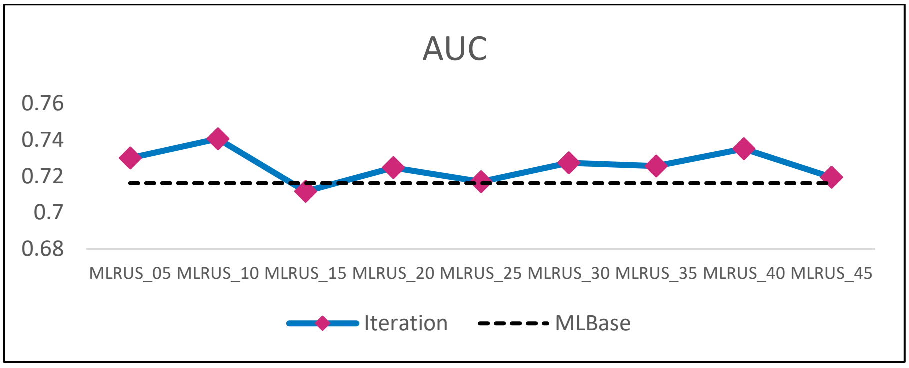


Figure 4-3: AUC trends Random Under Sampling

4.5.2 Random Over Sampling

Similar to the previous method, random over sampling was implemented iteratively. In this scenario, the full population belonging to the majority class (non-default customers) were considered and the minority class (default customers) were oversampled so that the dataset moved closer to being balanced at each iteration. Random over sampling works by taking all instances from the minority class and randomly duplicating these instances resulting in a higher number of observations for training and validation. The first iteration over-sampled a percentage of the customers from the minority class and considered all customers in the majority class such that the ratio of non-default to default customers became 95:05. Table 4.4 illustrates the characteristics of the training and validation dataset before and after each sampling iteration.

	# Records	**Non-Default**	**Default**	**Non-Default/Default Percentage**
Initial	117,189	115,223	1,966	98:02
Iteration 1	121,287	115,223	6,064	95:05
Iteration 2	128,025	115,223	12,802	90:10
Iteration 3	135,556	115,223	20,333	85:15
Iteration 4	144,028	115,223	28,805	80:20
Iteration 5	153,630	115,223	38,407	75:25
Iteration 6	164,604	115,223	49,381	70:30
Iteration 7	177,266	115,223	62,043	65:35
Iteration 8	192,056	115,223	76,833	60:40
Iteration 9	209,496	115,223	94,273	55:45

Table 4-4: Random Over Sampling

Table 4.4 highlights the increase in the default customers at each iteration. Three models were trained and tested at each iteration and the average AUC, recall and average class accuracy was recorded for each iteration. Table 4.4 details the results obtained for each of the three performance measures when the model was run using the test dataset where $\mathrm{ML}_{\mathrm{ROS\_(x)}}$ represents the results of the model which under samples the majority class such that the ratio of non-default to default is $100 - x{:}\,x$. Due to the increasing size in data at each iteration and the computational power required to run a model on data of that magnitude, the iterative process stopped at iteration 5 as results began to plateau. Overall there was considerable improvements across all performance measure at each iteration with a marginal increase in AUC to 0.7296 from 0.7161. Of all iterations, $\mathrm{ML}_{\mathrm{ROS\_25}}$ (over sampling minority such that the ratio of non-default to default is 75:25) outperformed the other sampling methods and the baseline model on both recall and average class accuracy. With regards AUC, there is no substantial change across all iterations. $\mathrm{ML}_{\mathrm{ROS\_25}}$ fitted to the training data with an average class accuracy of 74% and 63% recall. Overfitting is not apparent when compared to the results from $\mathrm{ML}_{\mathrm{ROS\_25}}$ on the unseen test set as highlighted in table 4.5. $\mathrm{ML}_{\mathrm{ROS\_25}}$ was compared with the optimal random under sampling method ($\mathrm{ML}_{\mathrm{RUS\_10}}$) to determine which technique would be brought forward to be used in the next experiment. This is outlined in section 4.5.3.

Model Name	Model Description	AUC	Recall	Average Class Accuracy
ML_{base}	Baseline	0.7161	58.5%	67.8%
ML_{ROS_05}	95-05	0.7219	59.7%	68.8%
ML_{ROS_10}	90-10	0.7296	60.9%	69.1%
ML_{ROS_15}	85-15	0.7280	60.9%	69.1%
ML_{ROS_20}	80-20	0.7247	60.1%	68.8%
ML_{ROS_25}	75-25	0.7278	61.3%	69.4%

Table 4-5: Results from Random Over Sampling

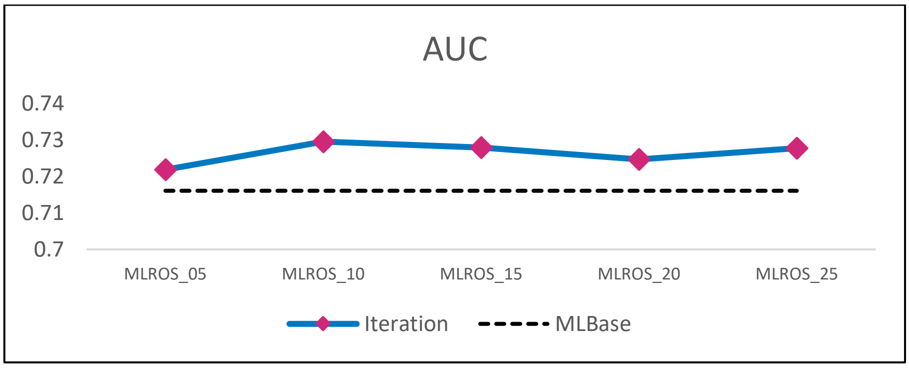


Figure 4-4: AUC trends Random Over Sampling

4.5.3 Model Comparisons

Table 4.6 illustrates the two models selected using random under sampling and random over sampling. Based on the performance measures ML_{RUS_10}, was the stronger of the two, in terms of AUC and recall. This means that ML_{RUS_10} was better at identifying the positive cases, identified as an important factor in section 3.7.1.

Model Name	Model Description	AUC	Recall	Average Class Accuracy
ML_{RUS_10}	90-10	0.7405	62.8%	69.4%
ML_{ROS_25}	75-25	0.7278	61.3%	69.4%

Table 4-6: Under/Over Sampling Model Comparison

To examine the differences between the two models statistically, a McNemar's test was applied to compare errors between the models for statistical significance. As discussed in section 2.9.3, McNemar’s test is a non-parametric statistical test which compares the disagreements between two sets of model predictions using a contingency table. The test determined that there was a statistically significant difference in the proportion of errors between the two models when executed on the test set with a 95% confidence interval (*p=0.003*).

	ML_{ROS_25} Correct	**ML_{ROS_25} Incorrect**
ML_{RUS_10} - Correct	21,827	948
ML_{RUS_10} - Incorrect	823	5,698

McNemar’s Test	
Chi-squared	8.6821
p-value	0.003

Table 4-7: McNemar's Test Sampling Models

Therefore, for the remaining experiments, the methods applied to create ML_{RUS_10} were used, i.e. the majority class was randomly under sampled in the training and validation datasets such that the ratio between non-default and default customers became 90:10.

4.6 Variable Selection

As discussed in section 2.7, variable selection is an important stage of model development. The research states the need to identify an effective subset of 10-20 variables (Hand & Henley, 1997) and also that when deciding variables for inclusion in a model, they should be logical, have a degree of predictive power and have a low correlation with each other (Anderson, 2007). Additionally, for this research, variable selection was beneficial in determining the most relevant transactional categories.

To identify important transactional variables, three variable selection techniques were experimented with, namely correlation variable selection, decision tree variable selection and random forest variable selection. The top 15 variables from each technique were selected and three models per technique were built utilising the top 5, top 10 and top 15 variables. These models were compared against the new baseline model ML_{RUS_10} discussed in the previous section.

The following experiments utilised the full population of customers, which was partitioned into a training and validation dataset and a test set as illustrated in table 4.8. The variable selection techniques were applied to the training/validation dataset.

	# Records	Non-Default	Default	Non-Default/Default Percentage
Train/Validation	234,377	230,445	3,932	98:02
Test	58,593	57,611	982	98:02
Total	**292,970**	**288,056**	**4,914**	**98:02**

Table 4-8: Variable Selection - Training and Test split

4.6.1 Correlation Analysis

Correlation matrices were developed to evaluate the inter-relationships between the independent transactional variables and also their relationship with the dependent variable. The Pearson correlation coefficient[14] was used to examine correlations. Due to the number of variables, it was not possible to create one correlation matrix for the full dataset. As an alternative, the matrices were built separately based on a number of categories and subsets of the data, e.g. monthly net spend, monthly spend features in the groceries category, all spend from 6 months previous.

Figures 4.5(a), (b), (c) and (d) illustrates the correlation matrices for all variables belonging to transactions which took place in the following categories respectively; Groceries, ATM Withdrawals, Dining and Auto. The blue cells in the figures below indicate variables which have a positive correlation whilst the red cells indicate those with a negative correlation. Non-correlated variables appear as white cells. Two independent variables with a Pearson correlation above 0.8 were considered for removal, with the variable having the strongest relationship with target kept as input variables into the predictive models.

[14] http://support.sas.com/documentation/cdl/en/procstat/63104/HTML/default/viewer.htm#procstat_corr_sect013.htm

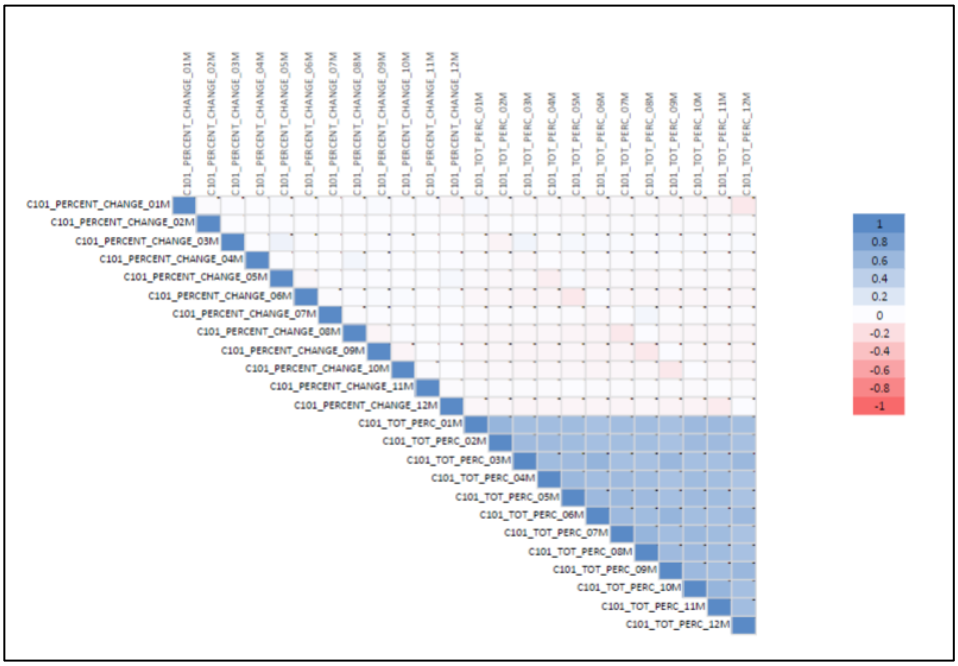


Figure 4-5(a): Grocery Category Correlation Matrix

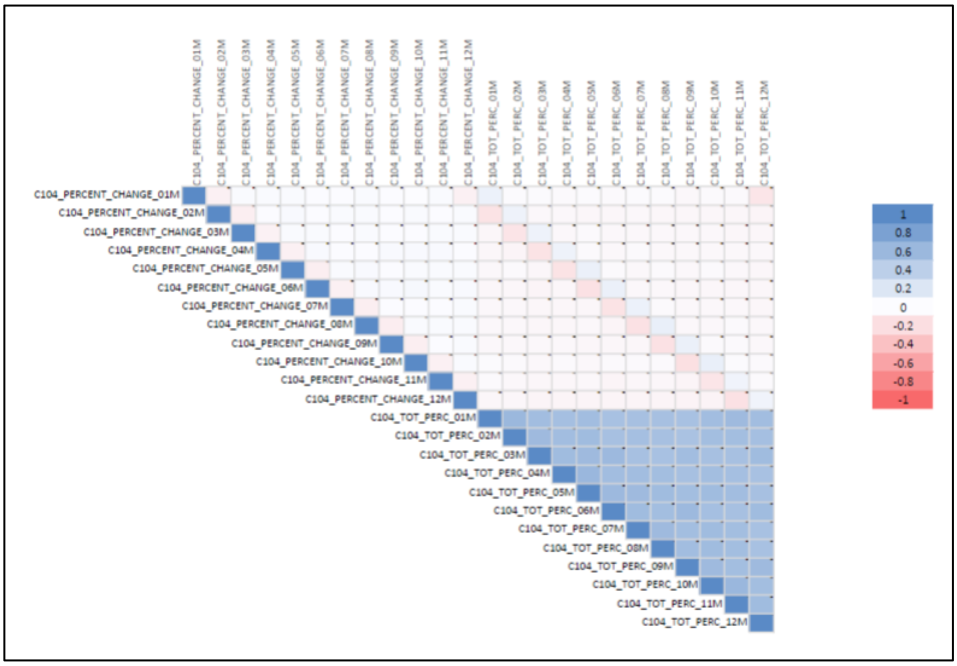


Figure 4-5(b): ATM Withdrawals Category Confusion Matrix

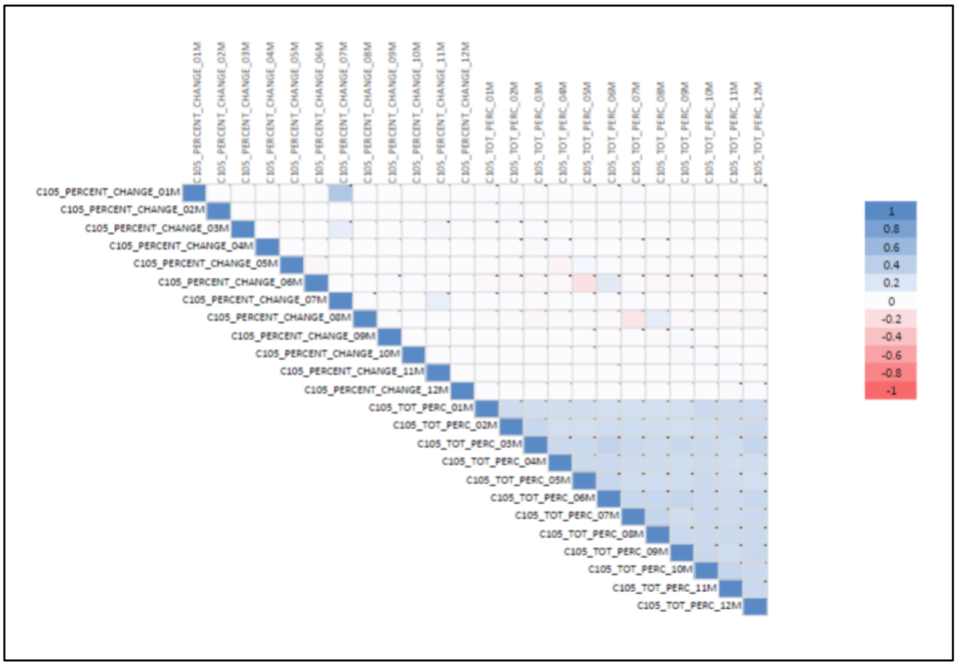


Figure 4-5(c): Dining Category Correlation Matrix

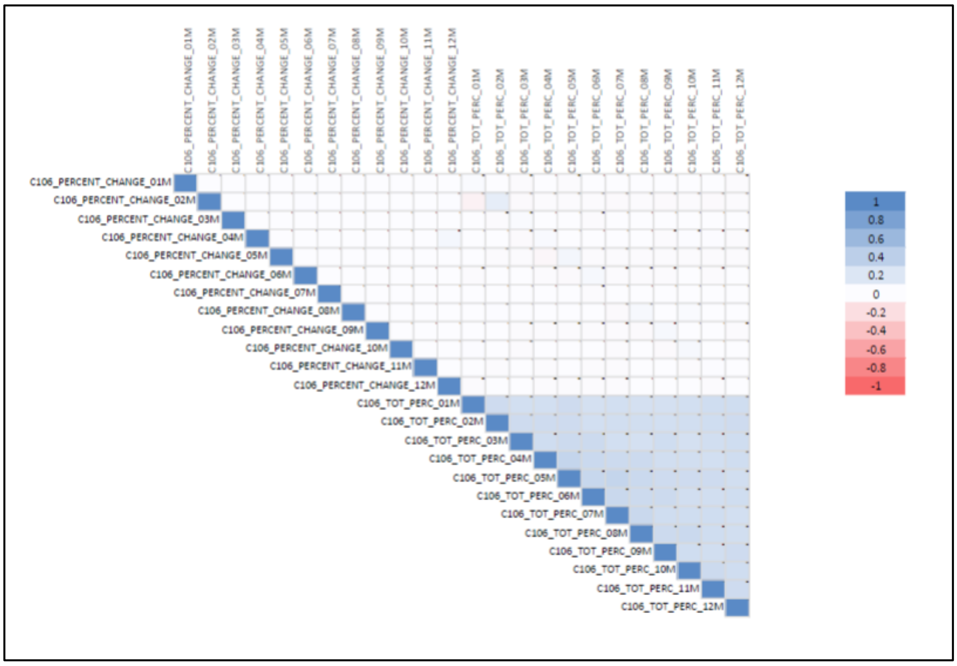


Figure 4-5(d): Auto Category Correlation Matrix

For the four categories visualised, there was a moderate positive correlation evident between some of the variables, particularly those variables outlining the total category spend as a percentage of the total spend on a monthly basis. However, none of these correlations exceeded the threshold of 0.8 so they were not considered for removal.

Similar analysis was undertaken for the remaining categories, details of which can be found in Appendix B.

In figure 4.6, a separate subset of variables was considered. This evaluated transactional variables that occurred in month 10 of the observation window (i.e. ten months prior to the observation point) across all categories. As illustrated in figure 4.6, there was no highly correlated variable evident. A similar process was undertaken for remaining eleven months with similar results. Details of these can be found in Appendix B.

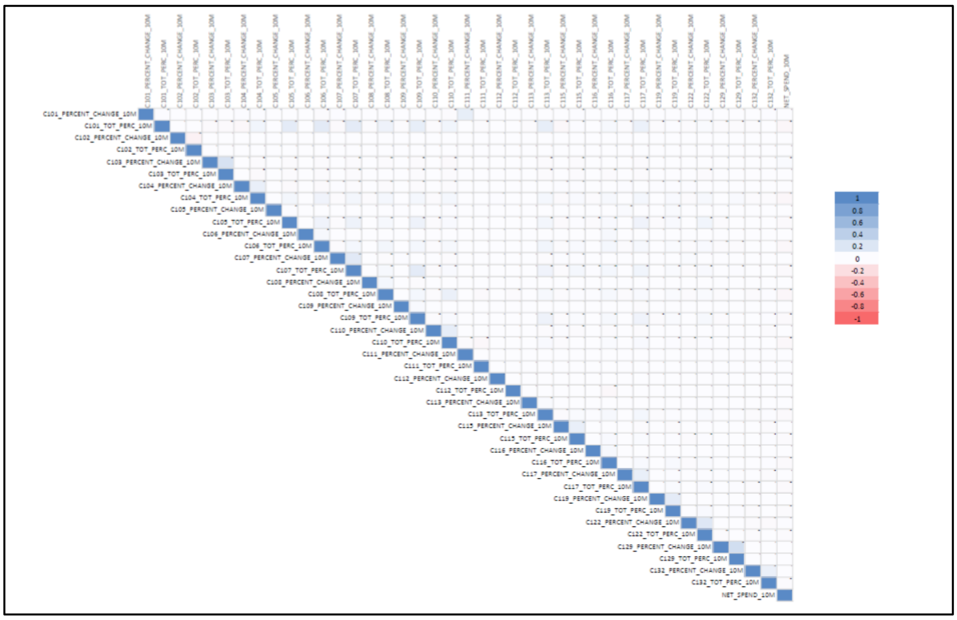


Figure 4-6: Month 10 Correlation Matrix

The final set of variables tested for correlations was all variables related to monthly net spend. As evidenced in figure 4.7, there was a strong positive correlation greater than 0.8 between all variables in this subset. The twelve variables' relationship with the target variable were examined and only one variable (NET_SPEND_01M) was retained to include as an input into a predictive model.

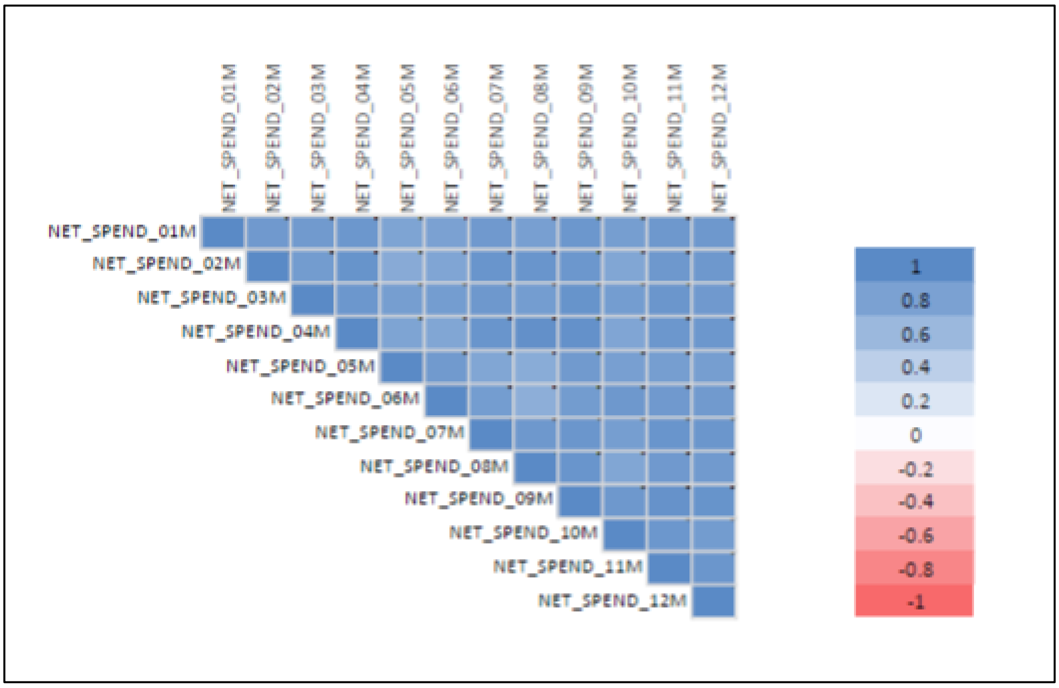


Figure 4-7: Net Spend Correlation Matrix

After the removal of highly correlated variables, the correlations between remaining variables and the target variable were calculated and the strongest 15 variables were considered for modelling purposes, details of which are available in figure 4.8.

Interestingly, the percentage spend on insurance on a monthly basis dominates the top 15 variables that had the strongest relationship with the target variable. The exploratory data analysis provided evidence that customers who have defaulted on their mortgage spend on average €100 more per month than those customers who have not defaulted on their mortgage. Additional categories that appeared in the top 15 highest correlated variables included ATM withdrawals, savings, gambling and health, with the latter three appearing in the top 5 variables. Similar to the insurance category, both ATM withdrawals and gambling spend were, on average, higher amongst the default customers. As discussed in the section 4.3.2, the non-default customers saved an average of €180 more than default customers per month with evidence of the latter withdrawing money from a savings account. This was a particularly interesting find with regards the identification of early warning indicators. The health category was the only category in the top 15 correlated variables where the average spend for both default and non-default was similar.

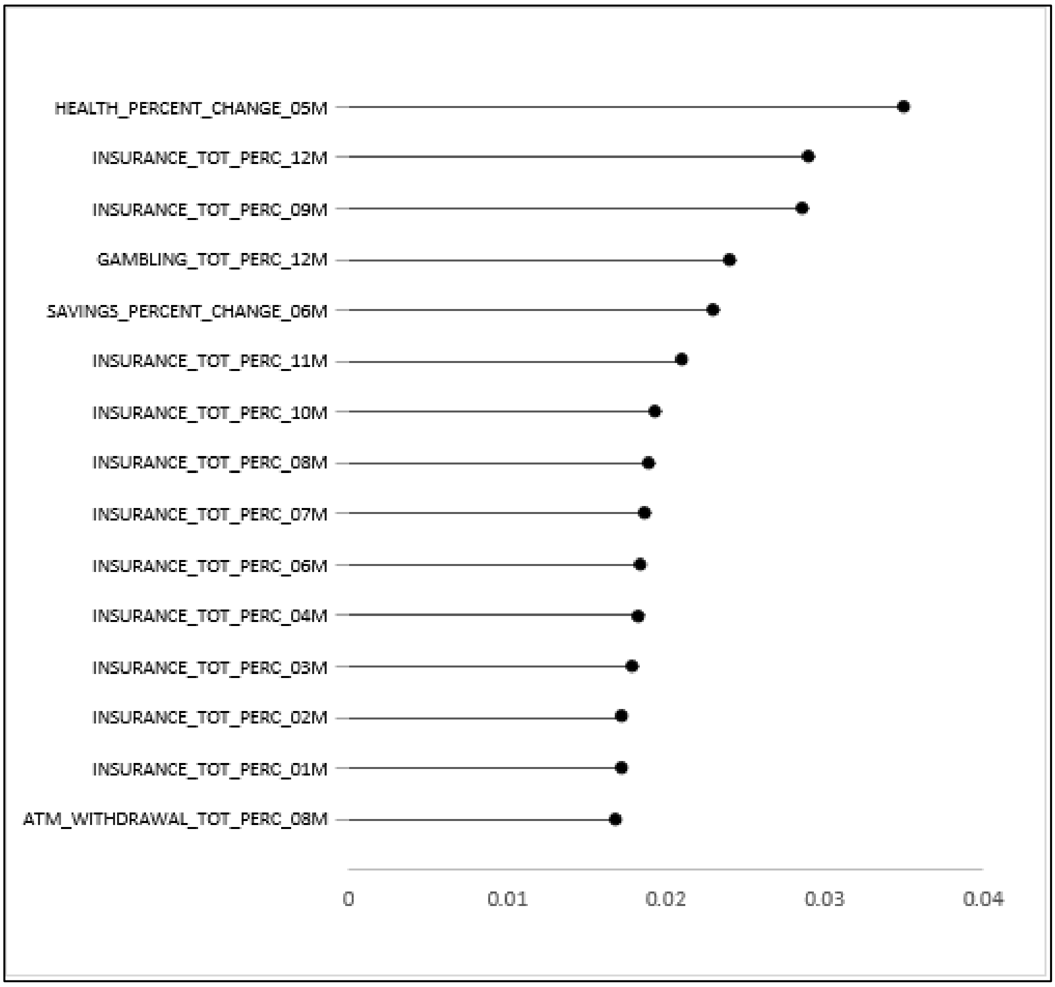


Figure 4-8: Top 15 Correlated Variables with Target

The top 15 transactional variables were split into groups containing the top 5, 10 and 15 variables and these were included in three separate models. As mentioned, these models were developed using the full population of customers as well as the random under sampling method chosen in section 4.5.3. The final training and validation dataset, after re-sampling and partitioning, is illustrated in table 4.9.

	# Records	Non-Default	Default	Non-Default/Default Percentage
Train	32,331	29,382	2,949	90:10
Validation	10,777	9,794	983	90:10
Total	**43,108**	**39,176**	**39,32**	**98:10**

Table 4-9: Variable Selection - Training and Validation split after Random Under Sampling

The baseline model, with just the original credit risk model variables, was re-trained on the full population and used to compare the outputs of the three models which included the top transactional variables based in their correlation with the target. Table 4.10 details the results produced for each performance measure when the models were executed using the test dataset, whilst figure 4.9 highlights the AUC for each model as extra features were added. It is evident from this figure that the performance starts to level out when more than 15 of the top features are added to the model.

		Train			Test		
Model Name	**Model Description**	**AUC**	**Recall**	**Avg Class Acc.**	**AUC**	**Recall**	**Avg Class Acc.**
ML_{RUS_10}	Baseline	0.7375	62.3%	70.1%	0.7207	60.1%	68.2%
ML_{COR_5}	Top 5	0.7454	63.4%	70.2%	0.7276	62.1%	68.4%
ML_{COR_10}	Top 10	0.7457	63.9%	70.2%	0.7284	61.9%	68.3%
ML_{COR_15}	Top 15	0.7465	64.9%	71.2%	0.7309	63.2%	68.9%

Table 4-10: Results including top Correlated Variables

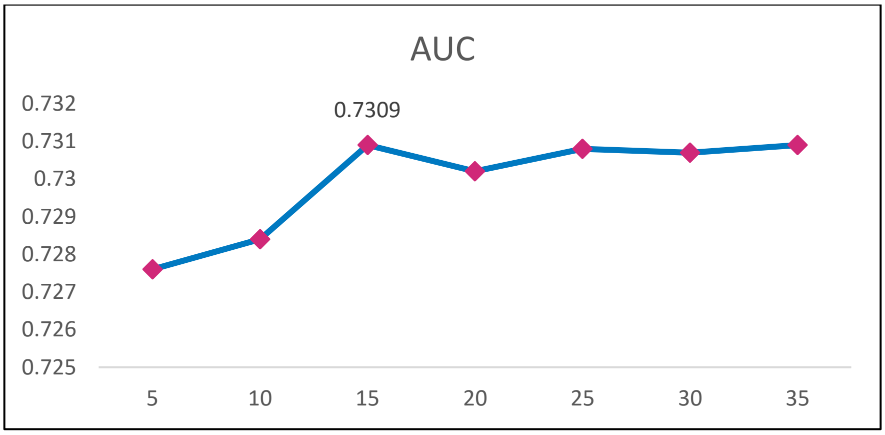


Figure 4-9: AUC trends with additional correlated features

Overall there was considerable improvements across all performance measures for each model with an increase in AUC to 0.7309 from 0.7206 on the test dataset. While the inclusion of additional features at each run did not improve the model considerably, ML_{COR_15} (inclusion of top 15 correlated variables) outperformed the other two models and the baseline model across all performance measures. ML_{COR_15} fitted to the training data with an average class accuracy of 71% and 65% recall. Therefore, overfitting was not apparent when compared to the results from ML_{COR_15} on the unseen test set as highlighted in table 4.10.

To examine, statistically, if there were differences between ML_{RUS_10} and ML_{COR_15}, a McNemar's test was applied. The test determined that there was a statistically significant difference in the proportion of errors between the two models when executed on the test set with a 95% confidence interval ($p < = 0.001$), as illustrated in Table 4-11.

	ML_{COR_15} Correct	**ML_{COR_15} Incorrect**
ML_{RUS_10} - Correct	42,151	2,423
ML_{RUS_10} - Incorrect	1,434	12,585

McNemar's Test	
Chi-squared	253.08
p-value	< = 0.001

Table 4-11: McNemar's Test ML_{RUS_10} V ML_{COR_15}

4.6.2 Decision Tree

Decision tree variable selection was undertaken on the dataset after the removal of independent correlated transactional variables. The process was developed using the HPSPLIT procedure on SAS which builds decision tree models and outputs metrics informing of variable importance. The importance measure is based on the change in the residual sum of squares at each split. The top 15 features with the highest relative importance were recorded and included in the prediction model, details of which can be found in figure 4.10.

Unlike the top 15 correlated variables, this method of variable selection produced a set of variables that span across a variety of the transactional categories. Some of the categories, such as gambling, insurance and ATM Withdrawals, were strong performers in both the decision tree variable selection method and the correlation selection method, highlighting their importance as potential early warning indicators. Interestingly, the variables that were selected based on their relative importance all occurred in the first 6 months of the observation window (i.e. features ending in 06M and 12M), reinforcing the decision to choose a 12-month observation window.

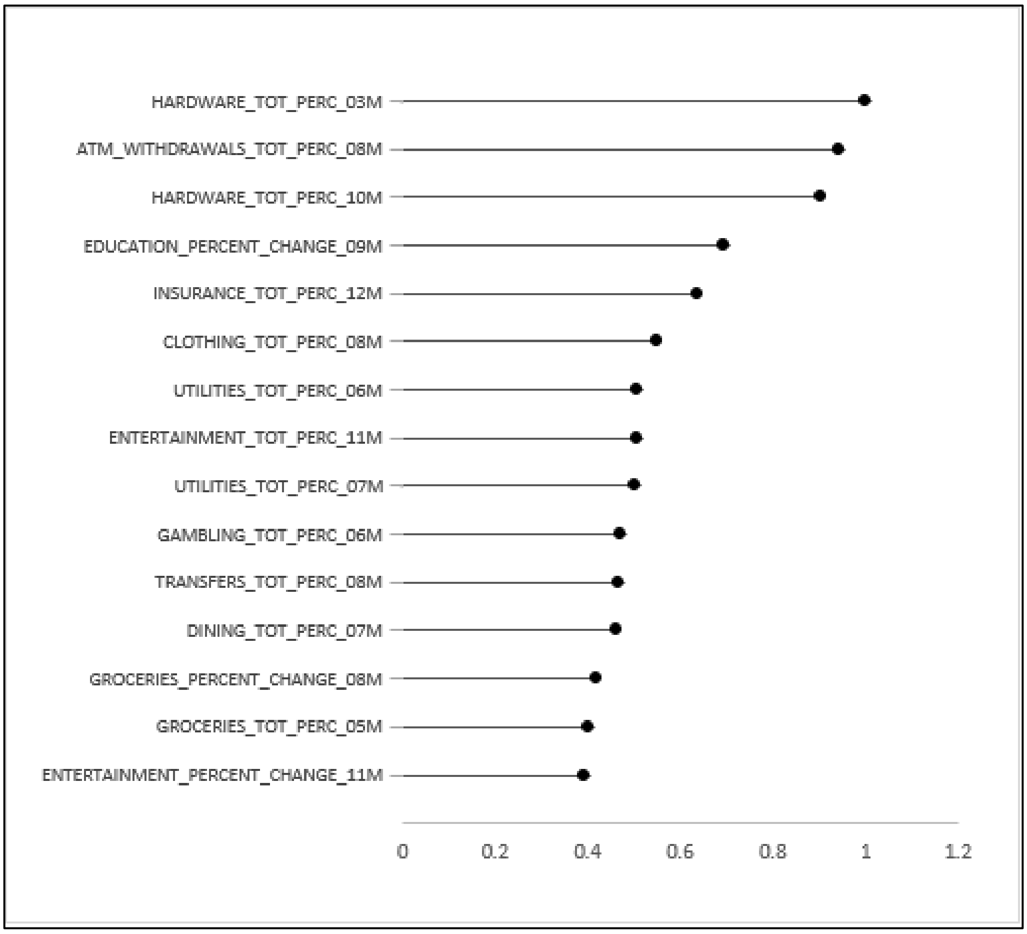


Figure 4-10: Top 15 HPSPLIT Variable Importance

While some of the spending in the categories outlined above could have been avoided or reduced, such as gambling, dining and entertainment, there are some categories which may have highlighted significant life events. For example, the variable, EDUCATION_PERCENT_CHANGE_09 could be related to a child starting back to school or college fees being due, whilst HARDWARE_TOT_PERC_03M and HARDWARE_TOT_PERC_10M could be related to ongoing home or garden renovations.

The top 15 transactional variables illustrated in figure 4.10 were split into groups containing the top 5, 10 and 15 variables and the baseline model (ML_{RUS_10}) was retrained three times with one group of variables included in each model. Table 4.12 details the results produced for each performance measure when the models were executed using the test dataset. Figure 4.11 highlights the AUC for each model as extra features were added. It is evident from this figure that the performance levels out when more than 15 of the top features are added to the model.

		Train			Test		
Model Name	**Model Description**	**AUC**	**Recall**	**Avg Class Acc.**	**AUC**	**Recall**	**Avg Class Acc.**
ML_{RUS_10}	Baseline	0.7375	62.3%	70.1%	0.7207	60.1%	68.2%
ML_{HPS_5}	Top 5	0.7408	64.1%	69.4%	0.7269	62.1%	69.0%
ML_{HPS_10}	Top 10	0.7432	63.5%	69.7%	0.7293	61.6%	68.6%
ML_{HPS_15}	Top 15	0.7414	65.7%	70.6%	0.7318	62.9%	69.1%

Table 4-12: Results including top HPSPLIT Variables

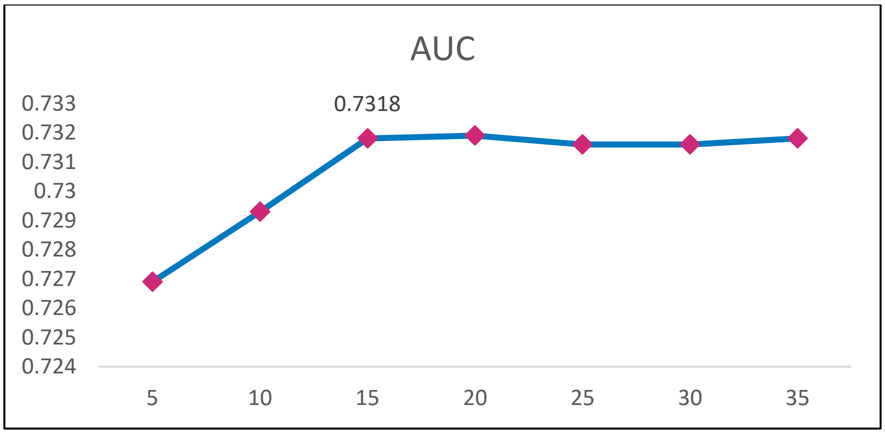


Figure 4-11: AUC trends with additional HPSPLIT features

Overall there was considerable improvements across all performance measures for each model with an increase in AUC to 0.7318 from 0.7206 on the test dataset. ML_{HPS_15} (inclusion of top 15 HPSPLIT variables) outperformed the other two models and the baseline model across all performance measures. ML_{HPS_15} fitted to the training data with an average class accuracy of 71% and 66% recall indicating that overfitting was not apparent when compared to the results from ML_{HPS_15} on the unseen test set as highlighted in table 4.12.

To examine, statistically, if there were differences between ML_{RUS_10} and ML_{HPS_15}, a McNemar's test was applied. The test determined that there was a statistically significant difference in the proportion of errors between the two models when executed on the test set with a 95% confidence interval ($p < = 0.001$). as illustrated in Table 4-13.

	ML_{HPS_15} Correct	**ML_{HPS_15} Incorrect**
ML_{RUS_10} - Correct	42,639	1,935
ML_{RUS_10} - Incorrect	1,333	12,686

McNemar's Test	
Chi-squared	104.58
p-value	< = 0.001

Table 4-13: McNemar's Test ML_{RUS_10} V ML_{HPS_15}

4.6.3 Random Forest

The final method used for variable selection was a random forest-based selection method. The process for this method was developed using the procedure HPFOREST on SAS which calculates variable importance by evaluating the loss reduction for each variable. This procedure was used on the training and validation dataset after the removal of highly correlated features and the top 15 variables were selected for modelling, details of which can be found in figure 4.12.

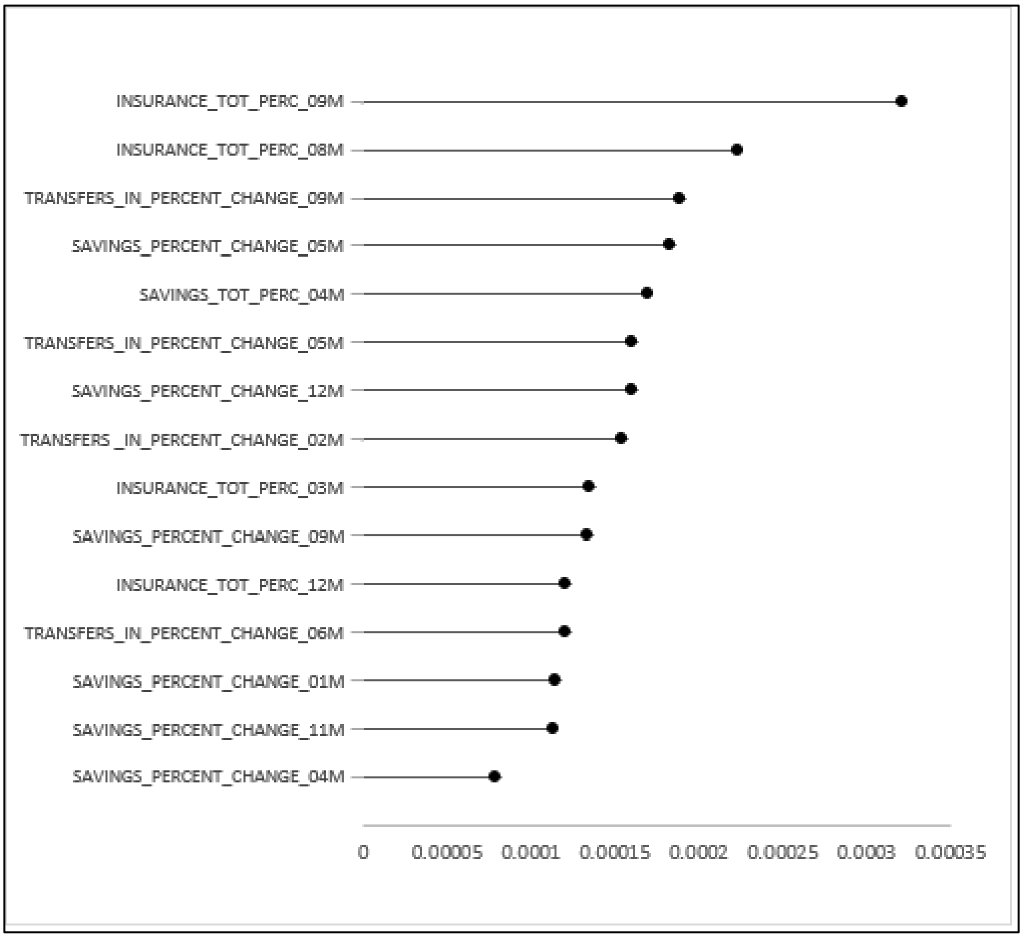


Figure 4-12: Top 15 HPFOREST Variable Importance

Similar to the correlation variable selection method, only a small number of transactional categories were represented in the top 15 variables produced as a result of the HPFOREST procedure, namely insurance, savings and transfers coming in. Based on the three variable selection technique, transactions taking place in the insurance category appear to be important given that it appeared consistently throughout each.

The top 15 transactional variables illustrated in figure 4.12 were split into groups containing the top 5, 10 and 15 variables and the baseline model (ML$_{\mathrm{RUS\_10}}$) was retrained three times with one group of variables included in each model. Table 4.13 details the results produced for each performance measure when the models were executed using the test dataset. Figure 4.13 highlights the AUC for each model as extra features were added. It is evident from this figure that the performance levels out when more than 15 of the top features are added to the model.

		Train			Test		
Model Name	**Model Description**	**AUC**	**Recall**	**Avg Class Acc.**	**AUC**	**Recall**	**Avg Class Acc.**
ML$_{\mathrm{RUS\_10}}$	Baseline	0.7375	62.3%	70.1%	0.7207	60.1%	68.2%
ML$_{\mathrm{HPF\_5}}$	Top 5	0.7444	64.1%	69.2%	0.7272	63.1%	68.9%
ML$_{\mathrm{HPF\_10}}$	Top 10	0.7454	64.6%	69.4%	0.7314	63.0%	68.7%
ML$_{\mathrm{HPF\_15}}$	Top 15	0.7460	65.2%	69.9%	0.7326	63.2%	68.8%

Table 4-14: Results including top HPFOREST Variables

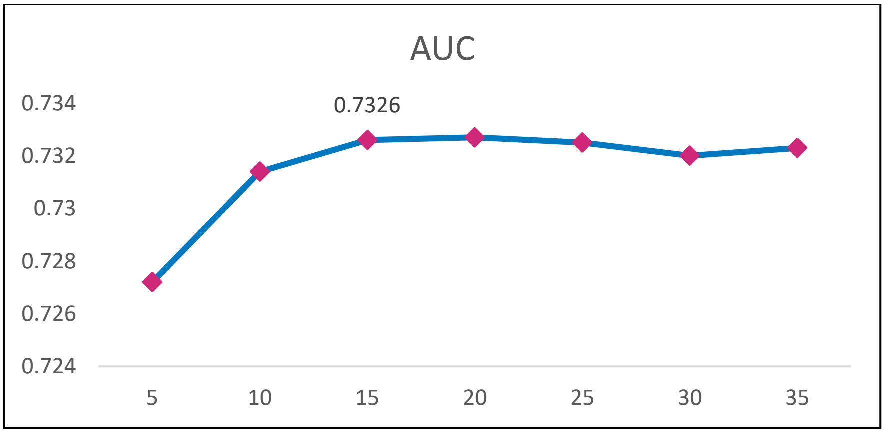


Figure 4-13: AUC trends with additional HPFOREST features

There were considerable improvements across all performance measures for each model with an increase in AUC to 0.7326 from 0.7206 on the test dataset. The models improved each time an additional five variables were added. ML$_{\mathrm{HPF\_15}}$ (inclusion of top 15 HPFOREST variables) outperformed the other two models and the baseline model across AUC and recall, with only marginal differences in the average class accuracy. ML_{HPF_15} fitted to the training data with an average class accuracy of 70% and 65% recall indicating that overfitting was not apparent when compared to the results from ML_{HPF_15} on the unseen test set as highlighted in table 4.14.

To examine, statistically, if there were differences between ML_{RUS_10} and ML_{HPF_15}, a McNemar's test was applied. The test determined that there was a statistically significant difference in the proportion of errors between the two models when executed on the test set with a 95% confidence interval (*p < = 0.001*), as illustrated in Table 4-15.

	ML_{HPF_15} Correct	**ML_{HPF_15} Incorrect**
ML_{RUS_10} - Correct	41,857	2.717
ML_{RUS_10} - Incorrect	1,647	12,372

McNemar's Test	
Chi-squared	261.68
p-value	< = 0.001

Table 4-15: McNemar's Test ML_{RUS_10} V ML_{HPF_15}

4.7 Interpretation of Results

Chapter 4 presented the implementation and evaluation of the experiments conducted throughout this research. The key objective of this research was to assess the use of non-traditional transactional features in predicting mortgage customers that will default on their repayment obligations. The baseline model was built using features from an existing credit risk model in Lender A. Throughout each experiment a holdout dataset was used for testing the performance of the models builds.

The baseline model was developed using a random under sampling method which removed a random sample of the majority class such that the dataset became fully balanced. An iterative process was implemented to determine an alternative methodology for dealing with the class imbalance which existed in the dataset. The baseline model was iteratively retrained using two methods of sampling; random under sampling and random over sampling. The expectation for this experiment was that the baseline model would be improved through the use of the two methods due to less, potentially important, data being excluded. A random under sampling method was chosen as the optimal strategy, increasing the AUC from 0.7161 to 0.7405, which excluded samples from the majority class such that there was a 90:10 split between the majority and minority class. While this method also involved removing data from the training dataset, it included eleven times more observations than the baseline model. The results from this experiment were in line with the expectation set out at the beginning.

To assess the transactional variables and identify those that were most important in terms of predicting whether mortgage customers would default or not, three variable selection techniques were applied. For the most part, the three selection methods identified a number of common transactional categories which was very beneficial in the identification of early warning indicators. For example, variables describing transactions which took place in the following categories; insurance, savings and gambling appeared in more than one of the variable selection techniques indicating their importance. Additionally, the month range for which these transactions took place was evident with 31 of the 45 variables selected relating to variables describing transactions that took place in the first six months of the observation window.

The baseline model was retrained including subsets of the variables selected throughout the variable selection process. The addition of these variables significantly improved the performance of the baseline model, with the inclusion of all 15 variables producing the best results for each selection method. Table 4.16 summarises the improvements observed. However, the differences in performance measure scores were marginal between the models developed with the top 5, 10 and 15 variables, asking the question if there is a need to continually add more variables.

Based on the study, it is recommended that Lender A utilise the important spend categories identified (i.e. savings, insurance and gambling) by either including them in their internal credit risk models or alternatively, developing suitable EWIs.

		Train			**Test**		
Model Name	**Model Description**	**AUC**	**Recall**	**Avg Class Acc.**	**AUC**	**Recall**	**Avg Class Acc.**
ML_{RUS_10}	Baseline	0.7375	62.3%	70.1%	0.7207	60.1%	68.2%
ML_{COR_15}	Top 15	0.7465	64.9%	71.2%	0.7309	63.2%	68.9%
ML_{HPS_15}	Top 15	0.7414	65.7%	70.6%	0.7318	62.9%	69.1%
ML_{HPF_15}	Top 15	0.7460	65.2%	69.9%	0.7326	63.2%	68.8%

Table 4-16: Summary of Variable Selection Models

5 CONCLUSION

5.1 Introduction

This chapter concludes the research paper, while also summarising the key findings from the research. The research question and objectives will be restated to serve as a reminder along with a discussion surrounding the contributions to the body of knowledge. An evaluation of the experiments, alongside an evaluation of the overall research will be demonstrated, including any limitations. Thoughts for future work and research will be discussed followed by concluding remarks.

5.2 Research Overview

The research carried out as part of the dissertation involved reviewing the literature in the field of mortgage arrears and default as well as the state of the art analytical approaches in the field of classification with a view to understanding the potential in the inclusion of non-traditional transactional features in an existing credit risk model. This review of the literature was used to design and implement a number of experiments to assess the predictive capability of transactional features for identifying customers who will default on their mortgage obligations. The following objectives were achieved:

- Review of the literature on mortgage arrears and default trends as well as best practices for credit scoring and predictive modelling
- Design and development of transactional features to be assessed for predicting mortgage arrears
- Design and build a baseline model using features available in existing credit risk model in Lender A that was used for comparisons
- Design of an experiment which compared sampling methods to overcome class imbalance
- Application of feature selection methods to the transactional features to identify most predictive
- Evaluation of the use of transactional in predicting customers who may default on their mortgage obligations.

The literature review revealed that the mortgage arrears crisis in Ireland was and is among the most severe experienced on record and although there has been a decreasing trend in the number of mortgages in default in the past four years, it still continues to cause distress to borrowers and vulnerabilities to lenders. The literature also revealed that one of the main factors associated with mortgage default is loan affordability, from which the level of disposable income is a driver of. Additionally, guidelines set out by the European Central Bank instructed financial institutions to adopt measures to further reduce and prevent loans defaulting, including the implementation and identification of Early Warning Indicators (EWIs).

The literature review on credit risk models revealed logistic regression as the industry standard approach to developing credit risk models due to its transparent nature, allowing financial institutions to provide relevant information to the regulator. Therefore, this study focused on utilising logistic regression and evaluating transactional level features to identify potential early warning indicators and establish if they would improve existing internal credit risk models.

5.3 Experimentation & Evaluation

In this research, logistic regression was used to evaluate the usefulness of non-traditional transactional features for predicting customers who may or may not default on their mortgage obligations. Default was defined as being 90 days past due i.e. three missed payments.

The target class was highly imbalanced with only 2% of the population classified as being in default. Two sampling methods were trialled to overcome the class imbalance problem; random under sampling and random over sampling. Approximately 270 million separate daily transactions were collected and, for each customer, were grouped into categories such as groceries, gambling and health. These were aggregated on a monthly basis and 491 features were developed based on the net spend percentage per category per month and the rate of change of net spend per category per month. Three feature selection methods were utilised to identify the top transactional features in terms of their usefulness in predicting mortgage default; correlation, decision tree-based feature selection and random forest-based feature selection. Nine further models were developed utilising subsets of the top transactional features. The three key performance metrics that were measured throughout this study were average class accuracy, AUC and recall. For statistical significance, McNemar's test was used to compare the errors between models generated with a significance level of 0.005.

The results from these experiments revealed that both under and over sampling models outperformed the baseline model, with random under sampling, such that the ratio of non-default to default customers was 90:10, achieving the highest AUC (0.7405 vs Baseline 0.7161), recall (62.8% vs Baseline 58.5%) and average class accuracy (69.4% vs Baseline 67.8%). Statistical analysis using a McNemar's test showed that there was a statistically significant difference in the results of the models.

The feature selection methods identified a number of transactional categories that proved beneficial in predicting mortgage default. The inclusion of this transactional level data resulted in the rate of success at predicting mortgage default to increase, indicating that the transactional variables are a useful determinant of mortgage default. The best results were produced using the top 15 features identified during the random forest feature selection method, achieving an AUC of 0.7326 vs the baseline AUC of 0.7202, recall of 63.2% vs the baseline recall of 60.1% and average class accuracy of 68.8% vs baseline of 68.2%. However, these results were only marginally better than the results produced for the two remaining feature selection techniques.

The results of these experiments will be useful for Lender A in two ways. Firstly, in early intervention by improving existing credit risk models and monitoring customer behaviour and secondly, as Lender A moves to develop EWIs, to conform with ECB guidance, the research in section 4.6 has shown some appropriate features from which to develop these indicators. Furthermore, these techniques identified a period of time Lender A should be focused on when examining EWIs, specifically the first 6 months prior to an event occurring.

Based on the analysis, it is recommended that Lender A include the following spend categories in future credit risk models, a) Insurance, b) Savings and c) Gambling, or utilise them to develop suitable EWIs; due to their importance across all three feature selection methods. The inclusion of these spend categories in future credit risk models or the development of EWIs will provide Lender A with additional foresight of customers who may default .

5.4 Limitations

A large portion of the data used for this study was based on transactional data from customers who had an open mortgage with Lender A. In a small number of cases, while a particular customer might have a mortgage account with Lender A, they might have a current account with a separate financial institution. It was not possible to collect transactional level data for these customers. While this is recognised as a potential limitation, it is uncommon for a customer to open a mortgage account and a current account in separate financial institutions. This holds true for Lender A, where approximately 80% of customers with a mortgage account, also hold a current account.

Additionally, due to the magnitude of the data, a full year of observation points was not made available by Lender A. Therefore, the data may not have fully represented changes in customer spending behaviours due to seasonality.

The results of this research may be affected due to regulations which apply to financial institutions, particularly with regards the use of low-level transactional data as used in this research. The introduction of the General Data Protection Regulation (GDPR) may also influence the results. The GDPR enforces direction surrounding data storage, ensuring companies only store data for a necessary period time. In addition, GDPR means that the processing of personal information and what it is used for is much more consent based which could pose as a significant limitation.

5.5 Summary of Contributions to Body of Knowledge

The following findings and results can be considered to be contributions to the body of knowledge achieved as part of this dissertation:

- Demonstrated that when handling imbalanced datasets, under sampling can be used to improve the performance
- Demonstrated that feature selection techniques such as correlation analysis, decision tree variable importance and random forest variable importance were successful in identifying EWIs
- Demonstrated that the transactional features created as part of the research were of predictive importance

5.6 Future Work and Recommendations

There are many additional experiments worth researching in this area. For example, this research focused on predicting if customers would default at any stage in the twelve months directly after the observation point. It would be interesting to experiment by shortening or expanding the length of the observation window or outcome window to see how that would impact the model performance and improve the capabilities of financial institutions.

Due to the current differences in the cost of living between urban and rural Ireland, it would be useful to introduce geographical features into the dataset to illustrate and evaluate the differences in customer spend at a county level. The use of GIS applications would be particularly useful for this analysis.

Finally, due to the limitation regarding data privacy and regulation on section 5.4, future work could take the form of considering and potentially amending policy to allow for transactional data to be used in order to prevent mortgage default going forward.

APPENDIX A

Feature List

Variable	Type	Description
CUSTOMER_RK	Interval	Customer Identifier
CUST_TRGT	Binary	Target Variable
TIME_SK	Interval	Observation Point Identifier
NET_SPEND_01M	Interval	Total Spend 1 month previous
C101_TOT_PERC_01M	Interval	Groceries Spend as a percentage of total spend 1 month previous
C101_TOT_PERC_02M	Interval	Groceries Spend as a percentage of total spend 2 months previous
C101_TOT_PERC_03M	Interval	Groceries Spend as a percentage of total spend 3 months previous
C101_TOT_PERC_04M	Interval	Groceries Spend as a percentage of total spend 4 months previous
C101_TOT_PERC_05M	Interval	Groceries Spend as a percentage of total spend 5 months previous
C101_TOT_PERC_06M	Interval	Groceries Spend as a percentage of total spend 6 months previous
C101_TOT_PERC_07M	Interval	Groceries Spend as a percentage of total spend 7 months previous
C101_TOT_PERC_08M	Interval	Groceries Spend as a percentage of total spend 8 months previous
C101_TOT_PERC_09M	Interval	Groceries Spend as a percentage of total spend 9 months previous
C101_TOT_PERC_10M	Interval	Groceries Spend as a percentage of total spend 10 months previous
C101_TOT_PERC_11M	Interval	Groceries Spend as a percentage of total spend 11 months previous
C101_TOT_PERC_12M	Interval	Groceries Spend as a percentage of total spend 12 months previous
C102_TOT_PERC_01M	Interval	Transfers In as a percentage of total spend 1 month previous
C102_TOT_PERC_02M	Interval	Transfers In as a percentage of total spend 2 months previous
C102_TOT_PERC_03M	Interval	Transfers In as a percentage of total spend 3 months previous
C102_TOT_PERC_04M	Interval	Transfers In as a percentage of total spend 4 months previous
C102_TOT_PERC_05M	Interval	Transfers In as a percentage of total spend 5 months previous
C102_TOT_PERC_06M	Interval	Transfers In as a percentage of total spend 6 months previous
C102_TOT_PERC_07M	Interval	Transfers In as a percentage of total spend 7 months previous
C102_TOT_PERC_08M	Interval	Transfers In as a percentage of total spend 8 months previous
C102_TOT_PERC_09M	Interval	Transfers In as a percentage of total spend 9 months previous
C102_TOT_PERC_10M	Interval	Transfers In as a percentage of total spend 10 months previous
C102_TOT_PERC_11M	Interval	Transfers In as a percentage of total spend 11 months previous
C102_TOT_PERC_12M	Interval	Transfers In as a percentage of total spend 12 months previous
C103_TOT_PERC_01M	Interval	Non ATM Withdrawals as a percentage of total spend 1 month previous
C103_TOT_PERC_02M	Interval	Non ATM Withdrawals as a percentage of total spend 2 months previous
C103_TOT_PERC_03M	Interval	Non ATM Withdrawals as a percentage of total spend 3 months previous
C103_TOT_PERC_04M	Interval	Non ATM Withdrawals as a percentage of total spend 4 months previous
C103_TOT_PERC_05M	Interval	Non ATM Withdrawals as a percentage of total spend 5 months previous
C103_TOT_PERC_06M	Interval	Non ATM Withdrawals as a percentage of total spend 6 months previous
C103_TOT_PERC_07M	Interval	Non ATM Withdrawals as a percentage of total spend 7 months previous
C103_TOT_PERC_08M	Interval	Non ATM Withdrawals as a percentage of total spend 8 months previous
C103_TOT_PERC_09M	Interval	Non ATM Withdrawals as a percentage of total spend 9 months previous
C103_TOT_PERC_10M	Interval	Non ATM Withdrawals as a percentage of total spend 10 months previous

C103_TOT_PERC_11M	Interval	Non ATM Withdrawals as a percentage of total spend 11 months previous
C103_TOT_PERC_12M	Interval	Non ATM Withdrawals as a percentage of total spend 12 months previous
C104_TOT_PERC_01M	Interval	ATM Withdrawals as a percentage of total spend 1 month previous
C104_TOT_PERC_02M	Interval	ATM Withdrawals as a percentage of total spend 2 months previous
C104_TOT_PERC_03M	Interval	ATM Withdrawals as a percentage of total spend 3 months previous
C104_TOT_PERC_04M	Interval	ATM Withdrawals as a percentage of total spend 4 months previous
C104_TOT_PERC_05M	Interval	ATM Withdrawals as a percentage of total spend 5 months previous
C104_TOT_PERC_06M	Interval	ATM Withdrawals as a percentage of total spend 6 months previous
C104_TOT_PERC_07M	Interval	ATM Withdrawals as a percentage of total spend 7 months previous
C104_TOT_PERC_08M	Interval	ATM Withdrawals as a percentage of total spend 8 months previous
C104_TOT_PERC_09M	Interval	ATM Withdrawals as a percentage of total spend 9 months previous
C104_TOT_PERC_10M	Interval	ATM Withdrawals as a percentage of total spend 10 months previous
C104_TOT_PERC_11M	Interval	ATM Withdrawals as a percentage of total spend 11 months previous
C104_TOT_PERC_12M	Interval	ATM Withdrawals as a percentage of total spend 12 months previous
C105_TOT_PERC_01M	Interval	Dining Spend as a percentage of total spend 1 month previous
C105_TOT_PERC_02M	Interval	Dining Spend as a percentage of total spend 2 months previous
C105_TOT_PERC_03M	Interval	Dining Spend as a percentage of total spend 3 months previous
C105_TOT_PERC_04M	Interval	Dining Spend as a percentage of total spend 4 months previous
C105_TOT_PERC_05M	Interval	Dining Spend as a percentage of total spend 5 months previous
C105_TOT_PERC_06M	Interval	Dining Spend as a percentage of total spend 6 months previous
C105_TOT_PERC_07M	Interval	Dining Spend as a percentage of total spend 7 months previous
C105_TOT_PERC_08M	Interval	Dining Spend as a percentage of total spend 8 months previous
C105_TOT_PERC_09M	Interval	Dining Spend as a percentage of total spend 9 months previous
C105_TOT_PERC_10M	Interval	Dining Spend as a percentage of total spend 10 months previous
C105_TOT_PERC_11M	Interval	Dining Spend as a percentage of total spend 11 months previous
C105_TOT_PERC_12M	Interval	Dining Spend as a percentage of total spend 12 months previous
C106_TOT_PERC_01M	Interval	Auto Spend as a percentage of total spend 1 month previous
C106_TOT_PERC_02M	Interval	Auto Spend as a percentage of total spend 2 months previous
C106_TOT_PERC_03M	Interval	Auto Spend as a percentage of total spend 3 months previous
C106_TOT_PERC_04M	Interval	Auto Spend as a percentage of total spend 4 months previous
C106_TOT_PERC_05M	Interval	Auto Spend as a percentage of total spend 5 months previous
C106_TOT_PERC_06M	Interval	Auto Spend as a percentage of total spend 6 months previous
C106_TOT_PERC_07M	Interval	Auto Spend as a percentage of total spend 7 months previous
C106_TOT_PERC_08M	Interval	Auto Spend as a percentage of total spend 8 months previous
C106_TOT_PERC_09M	Interval	Auto Spend as a percentage of total spend 9 months previous
C106_TOT_PERC_10M	Interval	Auto Spend as a percentage of total spend 10 months previous
C106_TOT_PERC_11M	Interval	Auto Spend as a percentage of total spend 11 months previous
C106_TOT_PERC_12M	Interval	Auto Spend as a percentage of total spend 12 months previous
C107_TOT_PERC_01M	Interval	Other Retail Spend as a percentage of total spend 1 month previous
C107_TOT_PERC_02M	Interval	Other Retail Spend as a percentage of total spend 2 months previous
C107_TOT_PERC_03M	Interval	Other Retail Spend as a percentage of total spend 3 months previous
C107_TOT_PERC_04M	Interval	Other Retail Spend as a percentage of total spend 4 months previous
C107_TOT_PERC_05M	Interval	Other Retail Spend as a percentage of total spend 5 months previous

C107_TOT_PERC_06M	Interval	Other Retail Spend as a percentage of total spend 6 months previous
C107_TOT_PERC_07M	Interval	Other Retail Spend as a percentage of total spend 7 months previous
C107_TOT_PERC_08M	Interval	Other Retail Spend as a percentage of total spend 8 months previous
C107_TOT_PERC_09M	Interval	Other Retail Spend as a percentage of total spend 9 months previous
C107_TOT_PERC_10M	Interval	Other Retail Spend as a percentage of total spend 10 months previous
C107_TOT_PERC_11M	Interval	Other Retail Spend as a percentage of total spend 11 months previous
C107_TOT_PERC_12M	Interval	Other Retail Spend as a percentage of total spend 12 months previous
C108_TOT_PERC_01M	Interval	Utilities Spend as a percentage of total spend 1 month previous
C108_TOT_PERC_02M	Interval	Utilities Spend as a percentage of total spend 2 months previous
C108_TOT_PERC_03M	Interval	Utilities Spend as a percentage of total spend 3 months previous
C108_TOT_PERC_04M	Interval	Utilities Spend as a percentage of total spend 4 months previous
C108_TOT_PERC_05M	Interval	Utilities Spend as a percentage of total spend 5 months previous
C108_TOT_PERC_06M	Interval	Utilities Spend as a percentage of total spend 6 months previous
C108_TOT_PERC_07M	Interval	Utilities Spend as a percentage of total spend 7 months previous
C108_TOT_PERC_08M	Interval	Utilities Spend as a percentage of total spend 8 months previous
C108_TOT_PERC_09M	Interval	Utilities Spend as a percentage of total spend 9 months previous
C108_TOT_PERC_10M	Interval	Utilities Spend as a percentage of total spend 10 months previous
C108_TOT_PERC_11M	Interval	Utilities Spend as a percentage of total spend 11 months previous
C108_TOT_PERC_12M	Interval	Utilities Spend as a percentage of total spend 12 months previous
C109_TOT_PERC_01M	Interval	Clothing Spend as a percentage of total spend 1 month previous
C109_TOT_PERC_02M	Interval	Clothing Spend as a percentage of total spend 2 months previous
C109_TOT_PERC_03M	Interval	Clothing Spend as a percentage of total spend 3 months previous
C109_TOT_PERC_04M	Interval	Clothing Spend as a percentage of total spend 4 months previous
C109_TOT_PERC_05M	Interval	Clothing Spend as a percentage of total spend 5 months previous
C109_TOT_PERC_06M	Interval	Clothing Spend as a percentage of total spend 6 months previous
C109_TOT_PERC_07M	Interval	Clothing Spend as a percentage of total spend 7 months previous
C109_TOT_PERC_08M	Interval	Clothing Spend as a percentage of total spend 8 months previous
C109_TOT_PERC_09M	Interval	Clothing Spend as a percentage of total spend 9 months previous
C109_TOT_PERC_10M	Interval	Clothing Spend as a percentage of total spend 10 months previous
C109_TOT_PERC_11M	Interval	Clothing Spend as a percentage of total spend 11 months previous
C109_TOT_PERC_12M	Interval	Clothing Spend as a percentage of total spend 12 months previous
C110_TOT_PERC_01M	Interval	Insurance Spend as a percentage of total spend 1 month previous
C110_TOT_PERC_02M	Interval	Insurance Spend as a percentage of total spend 2 months previous
C110_TOT_PERC_03M	Interval	Insurance Spend as a percentage of total spend 3 months previous
C110_TOT_PERC_04M	Interval	Insurance Spend as a percentage of total spend 4 months previous
C110_TOT_PERC_05M	Interval	Insurance Spend as a percentage of total spend 5 months previous
C110_TOT_PERC_06M	Interval	Insurance Spend as a percentage of total spend 6 months previous
C110_TOT_PERC_07M	Interval	Insurance Spend as a percentage of total spend 7 months previous
C110_TOT_PERC_08M	Interval	Insurance Spend as a percentage of total spend 8 months previous
C110_TOT_PERC_09M	Interval	Insurance Spend as a percentage of total spend 9 months previous
C110_TOT_PERC_10M	Interval	Insurance Spend as a percentage of total spend 10 months previous
C110_TOT_PERC_11M	Interval	Insurance Spend as a percentage of total spend 11 months previous
C110_TOT_PERC_12M	Interval	Insurance Spend as a percentage of total spend 12 months previous

C112_TOT_PERC_01M	Interval	Savings as a percentage of total spend 1 month previous
C112_TOT_PERC_02M	Interval	Savings as a percentage of total spend 2 months previous
C112_TOT_PERC_03M	Interval	Savings as a percentage of total spend 3 months previous
C112_TOT_PERC_04M	Interval	Savings as a percentage of total spend 4 months previous
C112_TOT_PERC_05M	Interval	Savings as a percentage of total spend 5 months previous
C112_TOT_PERC_06M	Interval	Savings as a percentage of total spend 6 months previous
C112_TOT_PERC_07M	Interval	Savings as a percentage of total spend 7 months previous
C112_TOT_PERC_08M	Interval	Savings as a percentage of total spend 8 months previous
C112_TOT_PERC_09M	Interval	Savings as a percentage of total spend 9 months previous
C112_TOT_PERC_10M	Interval	Savings as a percentage of total spend 10 months previous
C112_TOT_PERC_11M	Interval	Savings as a percentage of total spend 11 months previous
C112_TOT_PERC_12M	Interval	Savings as a percentage of total spend 12 months previous
C113_TOT_PERC_01M	Interval	Health Spend as a percentage of total spend 1 month previous
C113_TOT_PERC_02M	Interval	Health Spend as a percentage of total spend 2 months previous
C113_TOT_PERC_03M	Interval	Health Spend as a percentage of total spend 3 months previous
C113_TOT_PERC_04M	Interval	Health Spend as a percentage of total spend 4 months previous
C113_TOT_PERC_05M	Interval	Health Spend as a percentage of total spend 5 months previous
C113_TOT_PERC_06M	Interval	Health Spend as a percentage of total spend 6 months previous
C113_TOT_PERC_07M	Interval	Health Spend as a percentage of total spend 7 months previous
C113_TOT_PERC_08M	Interval	Health Spend as a percentage of total spend 8 months previous
C113_TOT_PERC_09M	Interval	Health Spend as a percentage of total spend 9 months previous
C113_TOT_PERC_10M	Interval	Health Spend as a percentage of total spend 10 months previous
C113_TOT_PERC_11M	Interval	Health Spend as a percentage of total spend 11 months previous
C113_TOT_PERC_12M	Interval	Health Spend as a percentage of total spend 12 months previous
C115_TOT_PERC_01M	Interval	Hardware Spend as a percentage of total spend 1 month previous
C115_TOT_PERC_02M	Interval	Hardware Spend as a percentage of total spend 2 months previous
C115_TOT_PERC_03M	Interval	Hardware Spend as a percentage of total spend 3 months previous
C115_TOT_PERC_04M	Interval	Hardware Spend as a percentage of total spend 4 months previous
C115_TOT_PERC_05M	Interval	Hardware Spend as a percentage of total spend 5 months previous
C115_TOT_PERC_06M	Interval	Hardware Spend as a percentage of total spend 6 months previous
C115_TOT_PERC_07M	Interval	Hardware Spend as a percentage of total spend 7 months previous
C115_TOT_PERC_08M	Interval	Hardware Spend as a percentage of total spend 8 months previous
C115_TOT_PERC_09M	Interval	Hardware Spend as a percentage of total spend 9 months previous
C115_TOT_PERC_10M	Interval	Hardware Spend as a percentage of total spend 10 months previous
C115_TOT_PERC_11M	Interval	Hardware Spend as a percentage of total spend 11 months previous
C115_TOT_PERC_12M	Interval	Hardware Spend as a percentage of total spend 12 months previous
C116_TOT_PERC_01M	Interval	Professional Services Spend as a percentage of total spend 1 month previous
C116_TOT_PERC_02M	Interval	Professional Services Spend as a percentage of total spend 2 months previous
C116_TOT_PERC_03M	Interval	Professional Services Spend as a percentage of total spend 3 months previous
C116_TOT_PERC_04M	Interval	Professional Services Spend as a percentage of total spend 4 months previous
C116_TOT_PERC_05M	Interval	Professional Services Spend as a percentage of total spend 5 months previous
C116_TOT_PERC_06M	Interval	Professional Services Spend as a percentage of total spend 6 months previous
C116_TOT_PERC_07M	Interval	Professional Services Spend as a percentage of total spend 7 months previous

C116_TOT_PERC_08M	Interval	Professional Services Spend as a percentage of total spend 8 months previous
C116_TOT_PERC_09M	Interval	Professional Services Spend as a percentage of total spend 9 months previous
C116_TOT_PERC_10M	Interval	Professional Services Spend as a percentage of total spend 10 months previous
C116_TOT_PERC_11M	Interval	Professional Services Spend as a percentage of total spend 11 months previous
C116_TOT_PERC_12M	Interval	Professional Services Spend as a percentage of total spend 12 months previous
C117_TOT_PERC_01M	Interval	Entertainment Spend as a percentage of total spend 1 month previous
C117_TOT_PERC_02M	Interval	Entertainment Spend as a percentage of total spend 2 months previous
C117_TOT_PERC_03M	Interval	Entertainment Spend as a percentage of total spend 3 months previous
C117_TOT_PERC_04M	Interval	Entertainment Spend as a percentage of total spend 4 months previous
C117_TOT_PERC_05M	Interval	Entertainment Spend as a percentage of total spend 5 months previous
C117_TOT_PERC_06M	Interval	Entertainment Spend as a percentage of total spend 6 months previous
C117_TOT_PERC_07M	Interval	Entertainment Spend as a percentage of total spend 7 months previous
C117_TOT_PERC_08M	Interval	Entertainment Spend as a percentage of total spend 8 months previous
C117_TOT_PERC_09M	Interval	Entertainment Spend as a percentage of total spend 9 months previous
C117_TOT_PERC_10M	Interval	Entertainment Spend as a percentage of total spend 10 months previous
C117_TOT_PERC_11M	Interval	Entertainment Spend as a percentage of total spend 11 months previous
C117_TOT_PERC_12M	Interval	Entertainment Spend as a percentage of total spend 12 months previous
C119_TOT_PERC_01M	Interval	Transport Spend as a percentage of total spend 1 month previous
C119_TOT_PERC_02M	Interval	Transport Spend as a percentage of total spend 2 months previous
C119_TOT_PERC_03M	Interval	Transport Spend as a percentage of total spend 3 months previous
C119_TOT_PERC_04M	Interval	Transport Spend as a percentage of total spend 4 months previous
C119_TOT_PERC_05M	Interval	Transport Spend as a percentage of total spend 5 months previous
C119_TOT_PERC_06M	Interval	Transport Spend as a percentage of total spend 6 months previous
C119_TOT_PERC_07M	Interval	Transport Spend as a percentage of total spend 7 months previous
C119_TOT_PERC_08M	Interval	Transport Spend as a percentage of total spend 8 months previous
C119_TOT_PERC_09M	Interval	Transport Spend as a percentage of total spend 9 months previous
C119_TOT_PERC_10M	Interval	Transport Spend as a percentage of total spend 10 months previous
C119_TOT_PERC_11M	Interval	Transport Spend as a percentage of total spend 11 months previous
C119_TOT_PERC_12M	Interval	Transport Spend as a percentage of total spend 12 months previous
C122_TOT_PERC_01M	Interval	Accommodation Spend as a percentage of total spend 1 month previous
C122_TOT_PERC_02M	Interval	Accommodation Spend as a percentage of total spend 2 months previous
C122_TOT_PERC_03M	Interval	Accommodation Spend as a percentage of total spend 3 months previous
C122_TOT_PERC_04M	Interval	Accommodation Spend as a percentage of total spend 4 months previous
C122_TOT_PERC_05M	Interval	Accommodation Spend as a percentage of total spend 5 months previous
C122_TOT_PERC_06M	Interval	Accommodation Spend as a percentage of total spend 6 months previous
C122_TOT_PERC_07M	Interval	Accommodation Spend as a percentage of total spend 7 months previous
C122_TOT_PERC_08M	Interval	Accommodation Spend as a percentage of total spend 8 months previous
C122_TOT_PERC_09M	Interval	Accommodation Spend as a percentage of total spend 9 months previous
C122_TOT_PERC_10M	Interval	Accommodation Spend as a percentage of total spend 10 months previous
C122_TOT_PERC_11M	Interval	Accommodation Spend as a percentage of total spend 11 months previous
C122_TOT_PERC_12M	Interval	Accommodation Spend as a percentage of total spend 12 months previous
C129_TOT_PERC_01M	Interval	Gambling Spend as a percentage of total spend 1 month previous
C129_TOT_PERC_02M	Interval	Gambling Spend as a percentage of total spend 2 months previous

C129_TOT_PERC_03M	Interval	Gambling Spend as a percentage of total spend 3 months previous
C129_TOT_PERC_04M	Interval	Gambling Spend as a percentage of total spend 4 months previous
C129_TOT_PERC_05M	Interval	Gambling Spend as a percentage of total spend 5 months previous
C129_TOT_PERC_06M	Interval	Gambling Spend as a percentage of total spend 6 months previous
C129_TOT_PERC_07M	Interval	Gambling Spend as a percentage of total spend 7 months previous
C129_TOT_PERC_08M	Interval	Gambling Spend as a percentage of total spend 8 months previous
C129_TOT_PERC_09M	Interval	Gambling Spend as a percentage of total spend 9 months previous
C129_TOT_PERC_10M	Interval	Gambling Spend as a percentage of total spend 10 months previous
C129_TOT_PERC_11M	Interval	Gambling Spend as a percentage of total spend 11 months previous
C129_TOT_PERC_12M	Interval	Gambling Spend as a percentage of total spend 12 months previous
C132_TOT_PERC_01M	Interval	Education Spend as a percentage of total spend 1 month previous
C132_TOT_PERC_02M	Interval	Education Spend as a percentage of total spend 2 months previous
C132_TOT_PERC_03M	Interval	Education Spend as a percentage of total spend 3 months previous
C132_TOT_PERC_04M	Interval	Education Spend as a percentage of total spend 4 months previous
C132_TOT_PERC_05M	Interval	Education Spend as a percentage of total spend 5 months previous
C132_TOT_PERC_06M	Interval	Education Spend as a percentage of total spend 6 months previous
C132_TOT_PERC_07M	Interval	Education Spend as a percentage of total spend 7 months previous
C132_TOT_PERC_08M	Interval	Education Spend as a percentage of total spend 8 months previous
C132_TOT_PERC_09M	Interval	Education Spend as a percentage of total spend 9 months previous
C132_TOT_PERC_10M	Interval	Education Spend as a percentage of total spend 10 months previous
C132_TOT_PERC_11M	Interval	Education Spend as a percentage of total spend 11 months previous
C132_TOT_PERC_12M	Interval	Education Spend as a percentage of total spend 12 months previous
C101_PERCENT_CHANGE_01M	Interval	Percent Change in Groceries Spend 1 month previous
C101_PERCENT_CHANGE_02M	Interval	Percent Change in Groceries Spend 2 months previous
C101_PERCENT_CHANGE_03M	Interval	Percent Change in Groceries Spend 3 months previous
C101_PERCENT_CHANGE_04M	Interval	Percent Change in Groceries Spend 4 months previous
C101_PERCENT_CHANGE_05M	Interval	Percent Change in Groceries Spend 5 months previous
C101_PERCENT_CHANGE_06M	Interval	Percent Change in Groceries Spend 6 months previous
C101_PERCENT_CHANGE_07M	Interval	Percent Change in Groceries Spend 7 months previous
C101_PERCENT_CHANGE_08M	Interval	Percent Change in Groceries Spend 8 months previous
C101_PERCENT_CHANGE_09M	Interval	Percent Change in Groceries Spend 9 months previous
C101_PERCENT_CHANGE_10M	Interval	Percent Change in Groceries Spend 10 months previous
C101_PERCENT_CHANGE_11M	Interval	Percent Change in Groceries Spend 11 months previous
C101_PERCENT_CHANGE_12M	Interval	Percent Change in Groceries Spend 12 months previous
C102_PERCENT_CHANGE_01M	Interval	Percent Change in Transfers In 1 month previous
C102_PERCENT_CHANGE_02M	Interval	Percent Change in Transfers In 2 months previous
C102_PERCENT_CHANGE_03M	Interval	Percent Change in Transfers In 3 months previous
C102_PERCENT_CHANGE_04M	Interval	Percent Change in Transfers In 4 months previous
C102_PERCENT_CHANGE_05M	Interval	Percent Change in Transfers In 5 months previous
C102_PERCENT_CHANGE_06M	Interval	Percent Change in Transfers In 6 months previous
C102_PERCENT_CHANGE_07M	Interval	Percent Change in Transfers In 7 months previous
C102_PERCENT_CHANGE_08M	Interval	Percent Change in Transfers In 8 months previous
C102_PERCENT_CHANGE_09M	Interval	Percent Change in Transfers In 9 months previous

C102_PERCENT_CHANGE_10M	Interval	Percent Change in Transfers In 10 months previous
C102_PERCENT_CHANGE_11M	Interval	Percent Change in Transfers In 11 months previous
C102_PERCENT_CHANGE_12M	Interval	Percent Change in Transfers In 12 months previous
C103_PERCENT_CHANGE_01M	Interval	Percent Change in Withdrawals 1 month previous
C103_PERCENT_CHANGE_02M	Interval	Percent Change in Withdrawals 2 months previous
C103_PERCENT_CHANGE_03M	Interval	Percent Change in Withdrawals 3 months previous
C103_PERCENT_CHANGE_04M	Interval	Percent Change in Withdrawals 4 months previous
C103_PERCENT_CHANGE_05M	Interval	Percent Change in Withdrawals 5 months previous
C103_PERCENT_CHANGE_06M	Interval	Percent Change in Withdrawals 6 months previous
C103_PERCENT_CHANGE_07M	Interval	Percent Change in Withdrawals 7 months previous
C103_PERCENT_CHANGE_08M	Interval	Percent Change in Withdrawals 8 months previous
C103_PERCENT_CHANGE_09M	Interval	Percent Change in Withdrawals 9 months previous
C103_PERCENT_CHANGE_10M	Interval	Percent Change in Withdrawals 10 months previous
C103_PERCENT_CHANGE_11M	Interval	Percent Change in Withdrawals 11 months previous
C103_PERCENT_CHANGE_12M	Interval	Percent Change in Withdrawals 12 months previous
C104_PERCENT_CHANGE_01M	Interval	Percent Change in ATM Withdrawals 1 month previous
C104_PERCENT_CHANGE_02M	Interval	Percent Change in ATM Withdrawals 2 months previous
C104_PERCENT_CHANGE_03M	Interval	Percent Change in ATM Withdrawals 3 months previous
C104_PERCENT_CHANGE_04M	Interval	Percent Change in ATM Withdrawals 4 months previous
C104_PERCENT_CHANGE_05M	Interval	Percent Change in ATM Withdrawals 5 months previous
C104_PERCENT_CHANGE_06M	Interval	Percent Change in ATM Withdrawals 6 months previous
C104_PERCENT_CHANGE_07M	Interval	Percent Change in ATM Withdrawals 7 months previous
C104_PERCENT_CHANGE_08M	Interval	Percent Change in ATM Withdrawals 8 months previous
C104_PERCENT_CHANGE_09M	Interval	Percent Change in ATM Withdrawals 9 months previous
C104_PERCENT_CHANGE_10M	Interval	Percent Change in ATM Withdrawals 10 months previous
C104_PERCENT_CHANGE_11M	Interval	Percent Change in ATM Withdrawals 11 months previous
C104_PERCENT_CHANGE_12M	Interval	Percent Change in ATM Withdrawals 12 months previous
C105_PERCENT_CHANGE_01M	Interval	Percent Change in Dining Spend 1 month previous
C105_PERCENT_CHANGE_02M	Interval	Percent Change in Dining Spend 2 months previous
C105_PERCENT_CHANGE_03M	Interval	Percent Change in Dining Spend 3 months previous
C105_PERCENT_CHANGE_04M	Interval	Percent Change in Dining Spend 4 months previous
C105_PERCENT_CHANGE_05M	Interval	Percent Change in Dining Spend 5 months previous
C105_PERCENT_CHANGE_06M	Interval	Percent Change in Dining Spend 6 months previous
C105_PERCENT_CHANGE_07M	Interval	Percent Change in Dining Spend 7 months previous
C105_PERCENT_CHANGE_08M	Interval	Percent Change in Dining Spend 8 months previous
C105_PERCENT_CHANGE_09M	Interval	Percent Change in Dining Spend 9 months previous
C105_PERCENT_CHANGE_10M	Interval	Percent Change in Dining Spend 10 months previous
C105_PERCENT_CHANGE_11M	Interval	Percent Change in Dining Spend 11 months previous
C105_PERCENT_CHANGE_12M	Interval	Percent Change in Dining Spend 12 months previous
C106_PERCENT_CHANGE_01M	Interval	Percent Change in Auto Spend 1 month previous
C106_PERCENT_CHANGE_02M	Interval	Percent Change in Auto Spend 2 months previous
C106_PERCENT_CHANGE_03M	Interval	Percent Change in Auto Spend 3 months previous
C106_PERCENT_CHANGE_04M	Interval	Percent Change in Auto Spend 4 months previous

C106_PERCENT_CHANGE_05M	Interval	Percent Change in Auto Spend 5 months previous
C106_PERCENT_CHANGE_06M	Interval	Percent Change in Auto Spend 6 months previous
C106_PERCENT_CHANGE_07M	Interval	Percent Change in Auto Spend 7 months previous
C106_PERCENT_CHANGE_08M	Interval	Percent Change in Auto Spend 8 months previous
C106_PERCENT_CHANGE_09M	Interval	Percent Change in Auto Spend 9 months previous
C106_PERCENT_CHANGE_10M	Interval	Percent Change in Auto Spend 10 months previous
C106_PERCENT_CHANGE_11M	Interval	Percent Change in Auto Spend 11 months previous
C106_PERCENT_CHANGE_12M	Interval	Percent Change in Auto Spend 12 months previous
C107_PERCENT_CHANGE_01M	Interval	Percent Change in Other Retail Spend 1 month previous
C107_PERCENT_CHANGE_02M	Interval	Percent Change in Other Retail Spend 2 months previous
C107_PERCENT_CHANGE_03M	Interval	Percent Change in Other Retail Spend 3 months previous
C107_PERCENT_CHANGE_04M	Interval	Percent Change in Other Retail Spend 4 months previous
C107_PERCENT_CHANGE_05M	Interval	Percent Change in Other Retail Spend 5 months previous
C107_PERCENT_CHANGE_06M	Interval	Percent Change in Other Retail Spend 6 months previous
C107_PERCENT_CHANGE_07M	Interval	Percent Change in Other Retail Spend 7 months previous
C107_PERCENT_CHANGE_08M	Interval	Percent Change in Other Retail Spend 8 months previous
C107_PERCENT_CHANGE_09M	Interval	Percent Change in Other Retail Spend 9 months previous
C107_PERCENT_CHANGE_10M	Interval	Percent Change in Other Retail Spend 10 months previous
C107_PERCENT_CHANGE_11M	Interval	Percent Change in Other Retail Spend 11 months previous
C107_PERCENT_CHANGE_12M	Interval	Percent Change in Other Retail Spend 12 months previous
C108_PERCENT_CHANGE_01M	Interval	Percent Change in Utilities Spend 1 month previous
C108_PERCENT_CHANGE_02M	Interval	Percent Change in Utilities Spend 2 months previous
C108_PERCENT_CHANGE_03M	Interval	Percent Change in Utilities Spend 3 months previous
C108_PERCENT_CHANGE_04M	Interval	Percent Change in Utilities Spend 4 months previous
C108_PERCENT_CHANGE_05M	Interval	Percent Change in Utilities Spend 5 months previous
C108_PERCENT_CHANGE_06M	Interval	Percent Change in Utilities Spend 6 months previous
C108_PERCENT_CHANGE_07M	Interval	Percent Change in Utilities Spend 7 months previous
C108_PERCENT_CHANGE_08M	Interval	Percent Change in Utilities Spend 8 months previous
C108_PERCENT_CHANGE_09M	Interval	Percent Change in Utilities Spend 9 months previous
C108_PERCENT_CHANGE_10M	Interval	Percent Change in Utilities Spend 10 months previous
C108_PERCENT_CHANGE_11M	Interval	Percent Change in Utilities Spend 11 months previous
C108_PERCENT_CHANGE_12M	Interval	Percent Change in Utilities Spend 12 months previous
C109_PERCENT_CHANGE_01M	Interval	Percent Change in Clothing Spend 1 month previous
C109_PERCENT_CHANGE_02M	Interval	Percent Change in Clothing Spend 2 months previous
C109_PERCENT_CHANGE_03M	Interval	Percent Change in Clothing Spend 3 months previous
C109_PERCENT_CHANGE_04M	Interval	Percent Change in Clothing Spend 4 months previous
C109_PERCENT_CHANGE_05M	Interval	Percent Change in Clothing Spend 5 months previous
C109_PERCENT_CHANGE_06M	Interval	Percent Change in Clothing Spend 6 months previous
C109_PERCENT_CHANGE_07M	Interval	Percent Change in Clothing Spend 7 months previous
C109_PERCENT_CHANGE_08M	Interval	Percent Change in Clothing Spend 8 months previous
C109_PERCENT_CHANGE_09M	Interval	Percent Change in Clothing Spend 9 months previous
C109_PERCENT_CHANGE_10M	Interval	Percent Change in Clothing Spend 10 months previous
C109_PERCENT_CHANGE_11M	Interval	Percent Change in Clothing Spend 11 months previous

C109_PERCENT_CHANGE_12M	Interval	Percent Change in Clothing Spend 12 months previous
C110_PERCENT_CHANGE_01M	Interval	Percent Change in Insurance Spend 1 month previous
C110_PERCENT_CHANGE_02M	Interval	Percent Change in Insurance Spend 2 months previous
C110_PERCENT_CHANGE_03M	Interval	Percent Change in Insurance Spend 3 months previous
C110_PERCENT_CHANGE_04M	Interval	Percent Change in Insurance Spend 4 months previous
C110_PERCENT_CHANGE_05M	Interval	Percent Change in Insurance Spend 5 months previous
C110_PERCENT_CHANGE_06M	Interval	Percent Change in Insurance Spend 6 months previous
C110_PERCENT_CHANGE_07M	Interval	Percent Change in Insurance Spend 7 months previous
C110_PERCENT_CHANGE_08M	Interval	Percent Change in Insurance Spend 8 months previous
C110_PERCENT_CHANGE_09M	Interval	Percent Change in Insurance Spend 9 months previous
C110_PERCENT_CHANGE_10M	Interval	Percent Change in Insurance Spend 10 months previous
C110_PERCENT_CHANGE_11M	Interval	Percent Change in Insurance Spend 11 months previous
C110_PERCENT_CHANGE_12M	Interval	Percent Change in Insurance Spend 12 months previous
C112_PERCENT_CHANGE_01M	Interval	Percent Change in Savings 1 month previous
C112_PERCENT_CHANGE_02M	Interval	Percent Change in Savings 2 months previous
C112_PERCENT_CHANGE_03M	Interval	Percent Change in Savings 3 months previous
C112_PERCENT_CHANGE_04M	Interval	Percent Change in Savings 4 months previous
C112_PERCENT_CHANGE_05M	Interval	Percent Change in Savings 5 months previous
C112_PERCENT_CHANGE_06M	Interval	Percent Change in Savings 6 months previous
C112_PERCENT_CHANGE_07M	Interval	Percent Change in Savings 7 months previous
C112_PERCENT_CHANGE_08M	Interval	Percent Change in Savings 8 months previous
C112_PERCENT_CHANGE_09M	Interval	Percent Change in Savings 9 months previous
C112_PERCENT_CHANGE_10M	Interval	Percent Change in Savings 10 months previous
C112_PERCENT_CHANGE_11M	Interval	Percent Change in Savings 11 months previous
C112_PERCENT_CHANGE_12M	Interval	Percent Change in Savings 12 months previous
C113_PERCENT_CHANGE_01M	Interval	Percent Change in Health Spend 1 month previous
C113_PERCENT_CHANGE_02M	Interval	Percent Change in Health Spend 2 months previous
C113_PERCENT_CHANGE_03M	Interval	Percent Change in Health Spend 3 months previous
C113_PERCENT_CHANGE_04M	Interval	Percent Change in Health Spend 4 months previous
C113_PERCENT_CHANGE_05M	Interval	Percent Change in Health Spend 5 months previous
C113_PERCENT_CHANGE_06M	Interval	Percent Change in Health Spend 6 months previous
C113_PERCENT_CHANGE_07M	Interval	Percent Change in Health Spend 7 months previous
C113_PERCENT_CHANGE_08M	Interval	Percent Change in Health Spend 8 months previous
C113_PERCENT_CHANGE_09M	Interval	Percent Change in Health Spend 9 months previous
C113_PERCENT_CHANGE_10M	Interval	Percent Change in Health Spend 10 months previous
C113_PERCENT_CHANGE_11M	Interval	Percent Change in Health Spend 11 months previous
C113_PERCENT_CHANGE_12M	Interval	Percent Change in Health Spend 12 months previous
C115_PERCENT_CHANGE_01M	Interval	Percent Change in Hardware Spend 1 month previous
C115_PERCENT_CHANGE_02M	Interval	Percent Change in Hardware Spend 2 months previous
C115_PERCENT_CHANGE_03M	Interval	Percent Change in Hardware Spend 3 months previous
C115_PERCENT_CHANGE_04M	Interval	Percent Change in Hardware Spend 4 months previous
C115_PERCENT_CHANGE_05M	Interval	Percent Change in Hardware Spend 5 months previous
C115_PERCENT_CHANGE_06M	Interval	Percent Change in Hardware Spend 6 months previous

C115_PERCENT_CHANGE_07M	Interval	Percent Change in Hardware Spend 7 months previous
C115_PERCENT_CHANGE_08M	Interval	Percent Change in Hardware Spend 8 months previous
C115_PERCENT_CHANGE_09M	Interval	Percent Change in Hardware Spend 9 months previous
C115_PERCENT_CHANGE_10M	Interval	Percent Change in Hardware Spend 10 months previous
C115_PERCENT_CHANGE_11M	Interval	Percent Change in Hardware Spend 11 months previous
C115_PERCENT_CHANGE_12M	Interval	Percent Change in Hardware Spend 12 months previous
C116_PERCENT_CHANGE_01M	Interval	Percent Change in Professional Services Spend 1 month previous
C116_PERCENT_CHANGE_02M	Interval	Percent Change in Professional Services Spend 2 months previous
C116_PERCENT_CHANGE_03M	Interval	Percent Change in Professional Services Spend 3 months previous
C116_PERCENT_CHANGE_04M	Interval	Percent Change in Professional Services Spend 4 months previous
C116_PERCENT_CHANGE_05M	Interval	Percent Change in Professional Services Spend 5 months previous
C116_PERCENT_CHANGE_06M	Interval	Percent Change in Professional Services Spend 6 months previous
C116_PERCENT_CHANGE_07M	Interval	Percent Change in Professional Services Spend 7 months previous
C116_PERCENT_CHANGE_08M	Interval	Percent Change in Professional Services Spend 8 months previous
C116_PERCENT_CHANGE_09M	Interval	Percent Change in Professional Services Spend 9 months previous
C116_PERCENT_CHANGE_10M	Interval	Percent Change in Professional Services Spend 10 months previous
C116_PERCENT_CHANGE_11M	Interval	Percent Change in Professional Services Spend 11 months previous
C116_PERCENT_CHANGE_12M	Interval	Percent Change in Professional Services Spend 12 months previous
C117_PERCENT_CHANGE_01M	Interval	Percent Change in Entertainment Spend 1 month previous
C117_PERCENT_CHANGE_02M	Interval	Percent Change in Entertainment Spend 2 months previous
C117_PERCENT_CHANGE_03M	Interval	Percent Change in Entertainment Spend 3 months previous
C117_PERCENT_CHANGE_04M	Interval	Percent Change in Entertainment Spend 4 months previous
C117_PERCENT_CHANGE_05M	Interval	Percent Change in Entertainment Spend 5 months previous
C117_PERCENT_CHANGE_06M	Interval	Percent Change in Entertainment Spend 6 months previous
C117_PERCENT_CHANGE_07M	Interval	Percent Change in Entertainment Spend 7 months previous
C117_PERCENT_CHANGE_08M	Interval	Percent Change in Entertainment Spend 8 months previous
C117_PERCENT_CHANGE_09M	Interval	Percent Change in Entertainment Spend 9 months previous
C117_PERCENT_CHANGE_10M	Interval	Percent Change in Entertainment Spend 10 months previous
C117_PERCENT_CHANGE_11M	Interval	Percent Change in Entertainment Spend 11 months previous
C117_PERCENT_CHANGE_12M	Interval	Percent Change in Entertainment Spend 12 months previous
C119_PERCENT_CHANGE_01M	Interval	Percent Change in Transport Spend 1 month previous
C119_PERCENT_CHANGE_02M	Interval	Percent Change in Transport Spend 2 months previous
C119_PERCENT_CHANGE_03M	Interval	Percent Change in Transport Spend 3 months previous
C119_PERCENT_CHANGE_04M	Interval	Percent Change in Transport Spend 4 months previous
C119_PERCENT_CHANGE_05M	Interval	Percent Change in Transport Spend 5 months previous
C119_PERCENT_CHANGE_06M	Interval	Percent Change in Transport Spend 6 months previous
C119_PERCENT_CHANGE_07M	Interval	Percent Change in Transport Spend 7 months previous
C119_PERCENT_CHANGE_08M	Interval	Percent Change in Transport Spend 8 months previous
C119_PERCENT_CHANGE_09M	Interval	Percent Change in Transport Spend 9 months previous
C119_PERCENT_CHANGE_10M	Interval	Percent Change in Transport Spend 10 months previous
C119_PERCENT_CHANGE_11M	Interval	Percent Change in Transport Spend 11 months previous
C119_PERCENT_CHANGE_12M	Interval	Percent Change in Transport Spend 12 months previous
C122_PERCENT_CHANGE_01M	Interval	Percent Change in Accommodation Spend 1 month previous

C122_PERCENT_CHANGE_02M	Interval	Percent Change in Accommodation Spend 2 months previous
C122_PERCENT_CHANGE_03M	Interval	Percent Change in Accommodation Spend 3 months previous
C122_PERCENT_CHANGE_04M	Interval	Percent Change in Accommodation Spend 4 months previous
C122_PERCENT_CHANGE_05M	Interval	Percent Change in Accommodation Spend 5 months previous
C122_PERCENT_CHANGE_06M	Interval	Percent Change in Accommodation Spend 6 months previous
C122_PERCENT_CHANGE_07M	Interval	Percent Change in Accommodation Spend 7 months previous
C122_PERCENT_CHANGE_08M	Interval	Percent Change in Accommodation Spend 8 months previous
C122_PERCENT_CHANGE_09M	Interval	Percent Change in Accommodation Spend 9 months previous
C122_PERCENT_CHANGE_10M	Interval	Percent Change in Accommodation Spend 10 months previous
C122_PERCENT_CHANGE_11M	Interval	Percent Change in Accommodation Spend 11 months previous
C122_PERCENT_CHANGE_12M	Interval	Percent Change in Accommodation Spend 12 months previous
C129_PERCENT_CHANGE_01M	Interval	Percent Change in Gambling Spend 1 month previous
C129_PERCENT_CHANGE_02M	Interval	Percent Change in Gambling Spend 2 months previous
C129_PERCENT_CHANGE_03M	Interval	Percent Change in Gambling Spend 3 months previous
C129_PERCENT_CHANGE_04M	Interval	Percent Change in Gambling Spend 4 months previous
C129_PERCENT_CHANGE_05M	Interval	Percent Change in Gambling Spend 5 months previous
C129_PERCENT_CHANGE_06M	Interval	Percent Change in Gambling Spend 6 months previous
C129_PERCENT_CHANGE_07M	Interval	Percent Change in Gambling Spend 7 months previous
C129_PERCENT_CHANGE_08M	Interval	Percent Change in Gambling Spend 8 months previous
C129_PERCENT_CHANGE_09M	Interval	Percent Change in Gambling Spend 9 months previous
C129_PERCENT_CHANGE_10M	Interval	Percent Change in Gambling Spend 10 months previous
C129_PERCENT_CHANGE_11M	Interval	Percent Change in Gambling Spend 11 months previous
C129_PERCENT_CHANGE_12M	Interval	Percent Change in Gambling Spend 12 months previous
C132_PERCENT_CHANGE_01M	Interval	Percent Change in Education Spend 1 month previous
C132_PERCENT_CHANGE_02M	Interval	Percent Change in Education Spend 2 months previous
C132_PERCENT_CHANGE_03M	Interval	Percent Change in Education Spend 3 months previous
C132_PERCENT_CHANGE_04M	Interval	Percent Change in Education Spend 4 months previous
C132_PERCENT_CHANGE_05M	Interval	Percent Change in Education Spend 5 months previous
C132_PERCENT_CHANGE_06M	Interval	Percent Change in Education Spend 6 months previous
C132_PERCENT_CHANGE_07M	Interval	Percent Change in Education Spend 7 months previous
C132_PERCENT_CHANGE_08M	Interval	Percent Change in Education Spend 8 months previous
C132_PERCENT_CHANGE_09M	Interval	Percent Change in Education Spend 9 months previous
C132_PERCENT_CHANGE_10M	Interval	Percent Change in Education Spend 10 months previous
C132_PERCENT_CHANGE_11M	Interval	Percent Change in Education Spend 11 months previous
C132_PERCENT_CHANGE_12M	Interval	Percent Change in Education Spend 12 months previous

APPENDIX B

Correlations

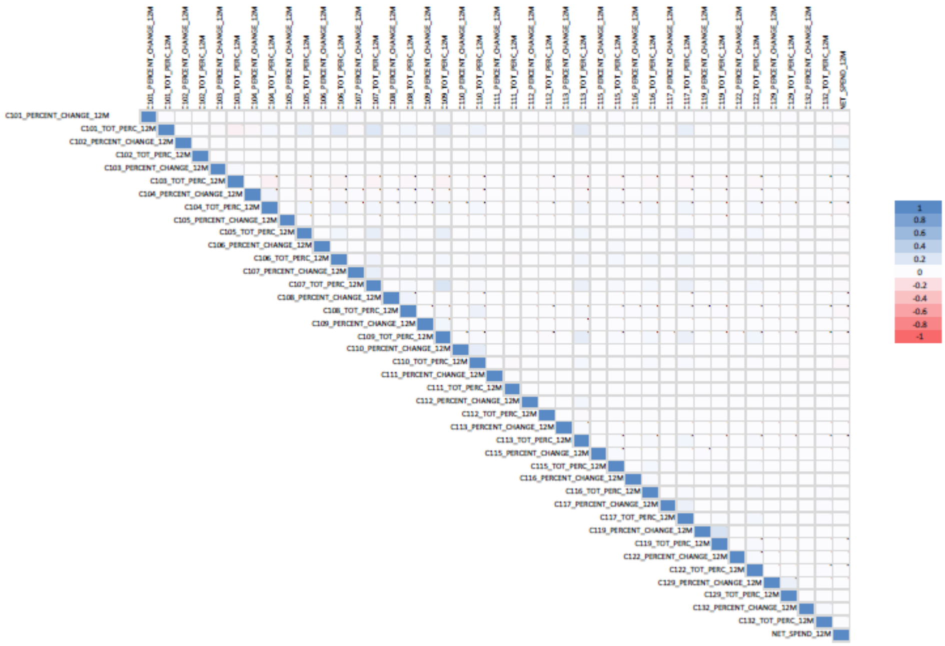


Figure B.1: Month 12 Correlation Matrix

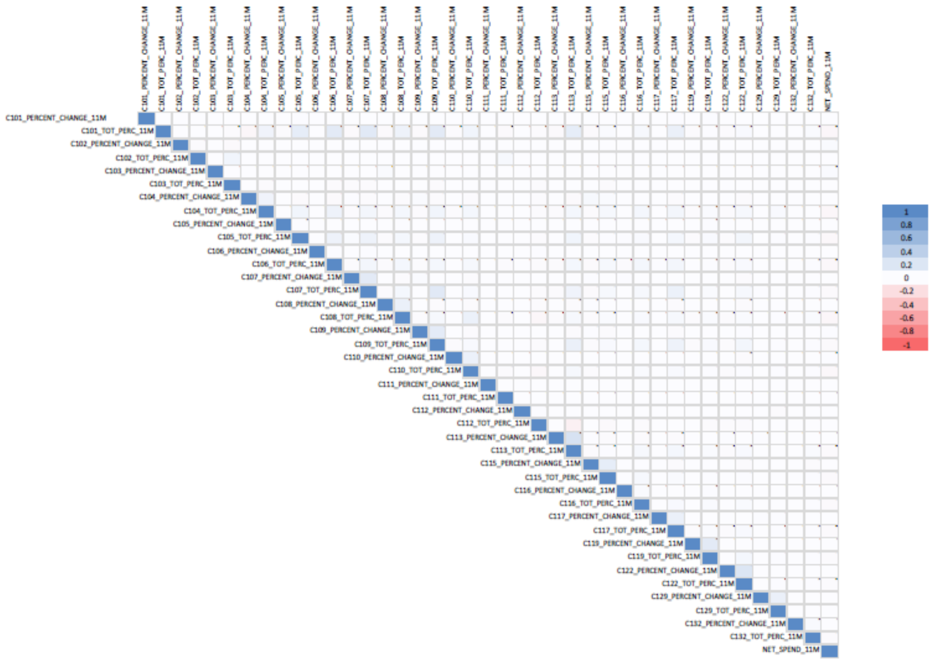


Figure B.2: Month 11 Correlation Matrix

BIBLIOGRPAHY

Anderson, R. (2007). The credit scoring toolkit: theory and practice for retail credit risk management and decision automation. Oxford: Oxford University Press.

Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. *Proceedings of the IADIS European conf. data mining*, 182–185. http://recipp.ipp.pt/handle/10400.22/136

Baesens, B., & Van Gestel, T. (2009). Credit Risk Management: Basic Concepts. Oxford University Press USA.

Baesens, B., Rosch, D., & Scheule, H. (2019). Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS. John Wiley & Sons.

Bolton, C. (2009). Logistic Regression and Its Application in Credit Scoring, Dissertation, University of Pretoria

Bradley, A. P. (1997), 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', Pattern recognition 30(7), 1145–1159.

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The Balanced Accuracy and Its Posterior Distribution. *2010 20th International Conference on Pattern Recognition*, 3121–3124. https://doi.org/10.1109/ICPR.2010.764

Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, *39*(3), 3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024

Chawla, N., 2010. Data mining for imbalanced datasets: An overview. In: Maimon, O., Rokach, L. (Eds.), Data Mining and Knowledge Discovery Handbook. Springer, US, pp. 875–886.

Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, P. (2002). SMOTE: Syhnthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research ,* (16)1, 321-357

Claessens, S., Dell'Ariccia, G., Igan, D., Laeven, L. (2010). Cross-country experiences and policy implications from the global financial crisis, *Economic Policy*, 25(62), 267–293 https://doi.org/10.1111/j.1468-0327.2010.00244.x

Connor, G., & Flavin, T. (2015). Strategic, unaffordability and dual-trigger default in the Irish mortgage market. *Journal of Housing Economics*, *28*, 59–75. https://doi.org/10.1016/j.jhe.2014.12.003

Dietterich, T.G. (1998). Approximate statistical test for computing supervised classification learning algorithms. *Neural Computation,* 10(7), 1895-1924

Dong, G., Lai, K. K., & Yen, J. (2010). Credit scorecard based on logistic regression with random coefficients. *Procedia Computer Science*, *1*(1), 2463–2468. https://doi.org/10.1016/j.procs.2010.04.278

Doughtery, E. R., Jianping, H., & Bittner, M. L. (2007). Validation of Computational Methods in Genomics. *Current Genomics*, *8*(1), 1–19. https://doi.org/10.2174/138920207780076956

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. and others (1996), Knowledge discovery and data mining: Towards a unifying framework., in 'KDD', Vol. 96, pp. 82–88. http://www.aaai.org/Papers/KDD/1996/KDD96-014

Feldman, D., & Gross, S. (2005). Mortgage Default: Classification Tree Analysis. *The Journal of Real Estate, Finance and Economics*, 30(4), 369-396. https://doi.org/10.1007/s11146-005-7013-7

Fitzpatrick, T., & Mues, C. (2016). An empirical comparison of classification algorithms for mortgage default prediction: evidence from a distressed mortgage market. *European Journal of Operational Research*, *249*(2), 427–439. https://doi.org/10.1016/j.ejor.2015.09.014

Frawley, W. J., Piatetsky-Shapiro, G. and Matheus, C. J. (1992), 'Knowledge discovery in databases: An overview', AI magazine 13(3), 57 http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/1011

Galindo, J., & Tamayo, P. (2000). Credit risk assessment using statistical and machine learning: Basic methodology and risk modelling applications. *Computational Economics,* 15(2), 107-143. https://doi.org/10.1023/A:1008699112516

Garcia, E., & He, H. (2009). Learning from Imbalances Data. *Transactions on Knowledge & Data Engineering*, 21(9), 1263-1284. https://doi.ieeecomputersociety.org/10.1109/TKDE.2008.239

Gerlach-Kristen, P., & Lyons, S. (2018). Determinants of mortgage arrears in Europe: evidence from household microdata. *International Journal of Housing Policy*, *18*(4), 545–567. https://doi.org/10.1080/19491247.2017.1357398

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research* 3, 1157–1182.

Hall, .M.A. (2000) Correlation-based feature selection of discrete and numeric class machine learning. *Proceedings of the Seventeenth International Conference on Machine Learning*, 359-366

Hand, D. J. (2001). Modelling consumer credit risk. *IMA Journal of Management Mathematics*, *12*(2), 139–155. https://doi.org/10.1093/imaman/12.2.139

Hand, D. J., & Henley, W. E. (1997). Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *160*(3), 523–541. https://doi.org/10.1111/j.1467-985X.1997.00078.x

Honohan, P. (2010). The Irish Banking Crisis: Regulatory and Financial Stability Policy 2003-2008. *Report for the Commission of Investigation into the Banking Sector in Ireland*

Hosmer, D., Lemeshow, S., 1989. Applied Logistic Regression. New York: John Wiley and Sons.

Japkowicz, N. (2000). The class imbalance problem: Significance and strategies. *Proceedings of the International Conference on Artificial Intelligence*.

Kelly, R., 2012. House Prices, Unemployment and Irish Mortgage Losses. *Research Technical Papers No. 10/RT/17, Central Bank of Ireland*

Kelly, R., & McCann, F. (2016). Some defaults are deeper than others: Understanding long-term mortgage arrears. *Journal of Banking & Finance*, *72*, 15–27. https://doi.org/10.1016/j.jbankfin.2016.07.006

Kelly, R., & O'Malley, T. (2016). The good, the bad and the impaired: A credit risk model of the Irish mortgage market. *Journal of Financial Stability*, *22*, 1–9. https://doi.org/10.1016/j.jfs.2015.09.005

Kennedy, K., Mac Namee, B., Delany, S. J., O'Sullivan, M., & Watson, N. (2013). A window of opportunity: Assessing behavioural scoring. *Expert Systems with Applications*, *40*(4), 1372–1380. https://doi.org/10.1016/j.eswa.2012.08.052

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

Kleinbaum, D. G., Klein, M., & Pryor, E. R. (2002). *Logistic regression: a self-learning text* (2nd ed). New York: Springer.

Kohavi, R., & Sommerfield, D. (1995). Feature subset selection using the wrapper model: Overfitting and dynamic search space topology. *The First International Conference on Knowledge Discovery and Data Mining,* 192– 197.

Koller, D. & Sahami, M. (1996). Toward Optimal Feature Selection. *Proceedings of the 13th International Conference on Machine Learning, ICML-96, Bari, Italy*, 284-292.

Kubat, M., & Matwin, S. (1997). Addressing the Curse of Imbalanced Training Sets: One Sided Selection. *Proceedings of the Fourteenth International Conference on Machine Learning*, 179-186.

Kumar, V. (2014). Feature Selection: A literature Review. *The Smart Computing Review*, *4*(3). https://doi.org/10.6029/smartcr.2014.03.007

Laeven, L., & Valencia, F. (2013). Systemic Banking Crises Database. *IMF Economic Review*, *61*(2), 225–270. https://doi.org/10.1057/imfer.2013.12

Ling, C., & Li, C. (1998). Data Mining for Direct Marketing Problems and Solutions. Proceedings *of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98) New York, NY. AAAI Press.*

Nyberg, P. (2011). Misjudging Risk: Causes of the Systemic Banking Crisis in Ireland. *Report of the Commission of Investigation into the Banking Sector in Ireland.*

McCann, F. (2017). Resolving a Non-Performing Loan crisis: The ongoing case of the Irish mortgage market. *Research Technical Papers No. 10/RT/17, Central Bank of Ireland*.

Regling, K. and Watson, M. (2010) A Preliminary Report on The Sources of Ireland's Banking Crisis, Dublin, Stationery Office.

Sánchez-Maroño, N., Alonso-Betanzos, A., & Tombilla-Sanromán, M. (2007). Filter Methods for Feature Selection – A Comparative Study. In H. Yin, P. Tino, E. Corchado, W. Byrne, & X. Yao (Eds.), *Intelligent Data Engineering and Automated Learning - IDEAL 2007* (Vol. 4881, pp. 178–187). https://doi.org/10.1007/978-3-540-77226-2_19

Shearer, C. (2000), 'The CRISP-DM model: the new blueprint for data mining', Journal of data warehousing 5(4), 13–22

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Thomas, L. (2009). Consumer credit models: pricing, profit and portfolios. Oxford University Press, USA.

Thomas, L. C. (2001). Time will tell: behavioural scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics*, *12*(1), 89–103. https://doi.org/10.1093/imaman/12.1.89

Whitley, J. K., Windram, R., & Cox, P. (2004). *An Empirical Model of Household Arrears*. *Bank of England Working Paper No, 214.* https://doi.org/10.2139/ssrn.598886