University of Canberra

Technology Capstone Research Project PG (11522)

Semester 1 2024

Final Report on

Developing a Data-Driven Home Loan Origination Scorecard: Enhancing Risk Management and Decision- Making in the Australian Mortgage Industry

Group Name: 11522-24S1-02

Submitted by:

Name	**Student ID**
Prahlad Panthi	U3248319
Shital Ghimire	U3218501
Aakash Tamang	U3248348
Birat Jung Thapa	U3245672

TABLE OF CONTENTS

TABLE OF FIGURES

EXECUTIVE SUMMARY

The Australian housing market is experiencing a concerning trend of increasing homeowner debt due to high inflation rates and sluggish salary growth. This poses a significant risk to the economy, as defaults on mortgage loans could lead to a substantial downturn. To better understand and address this issue, our project focused on analysing the factors influencing the home loan industry in Australia and other nations.

We conducted a comprehensive review of articles, case studies, and literature to gain insights into the complexities of the home loan market. Leveraging Machine Learning techniques, particularly utilizing the SAS Viya platform and the HMEQ dataset, we aimed to develop a predictive model to assess the likelihood of loan default.

Throughout the project, we encountered various challenges, including project initiation, task planning, workload distribution, outcome evaluation, integrating the model for wider use, and communication management. To address these challenges, we adopted the Agile Methodology for Project Management, ensuring efficient coordination and execution of tasks.

Ultimately, we successfully developed a predictive model capable of assessing loan default risk based on credit history and financial information provided by applicants. This achievement not only contributes to addressing the challenges of the Australian housing market but also provides valuable experience in project management techniques and data analytics. Overall, our project has resulted in personal and professional growth, equipping us with valuable skills and insights for future endeavours.

INTRODUCTION

This project aims to contribute to decision making and risk management in the Australian mortgage market through the development of advanced home origination models and delinquency prediction algorithms. By leveraging data driven insights, this project seeks to enhance the efficiency, accuracy, and sustainability of lending practices while addressing evolving market dynamics and regulatory requirements. Throughout this project, significant achievements have been made in the development and evaluation of sophisticated models.

The Australian mortgage market has been a critical component of the nation's economy, serving as a primary avenue for homeownership and property investment. Over the years, this market has witnessed significant fluctuations and transformation, influenced by various economic, regulatory, and societal factors. [1]

In recent times, the market dynamics have been shaped by events such as Global Financial Crisis (GFC), regulatory reforms, technological advancements, and changing consumer preferences. These developments have underscored the importance of robust decision-making frameworks, risk management strategies, and innovative solutions within the mortgage industry. [1]

Responsible lending is a critical aspect of managing credit and debt globally. While access to credit can assist consumers in spreading out the cost of expensive purchases or handling unexpected expenses, it's crucial to recognize the potential consequences if repayments become unmanageable [2]. Overtime, Borrowers and lenders has become more cautious due to economic uncertainty, leading to stricter lending standards [1].

Against this backdrop, our project seeks to explore the landscape of the Australian mortgage market and examine the home origination models and delinquency prediction in facilitating lending practices. By evaluating the market trend, analysing the loan product, we aim to provide insights into the dynamics of the mortgage market and its implications for lenders, borrowers, and regulators. Our primary focus remains in developing home loan originations models that accurately predict borrower creditworthiness, analyse delinquency prediction algorithms to identify key risk factors and variables influencing loan defaults and provide actionable insights and recommendations to banks and financial institutions for improving lending practices and mitigating risks.

PROJECT PERSONNEL

Role	Name	Contact
Sponsors	Allan Tham	allan.tham@sas.com
	David Bremstaller	David.Bremstaller@sas.com
	Lucy Biasi	Lucy.Biasi@sas.com
Project Mentor	Jeanette Cotterill	jeanette.cotterill@canberra.edu.au
Project Lead	Shital Ghimire	u3218501@uni.canberra.edu.au
Product Lead	Aakash Tamang	u3248348@uni.canberra.edu.au
Technical Lead	Birat Jung Thapa	u3245672@uni.canberra.edu.au
Communication Lead	Prahlad Panthi	u3248319@uni.canberra.edu.au

PROBLEM STATEMENT

The Australian mortgage market has grown highly competitive, resulting in aggressive lending approaches and potential hazards for both banks and borrowers. In the quest to acquire and retain customers, banks may unintentionally commit to riskier clients at narrow profit margins, impacting their future financial viability. Moreover, the existing environment for home loans in Australia, marked by escalating interest rates, increased funding expenses, and the impending fixed-rate cliff, poses distinctive obstacles for both lenders and borrowers.

To address these complex dynamics and mitigate potential risks, it is imperative to establish data-driven models for home loan origination, segmentation, and other analytical outputs. By addressing this challenge, our goal is to improve decision-making, risk management, and overall efficiency within the mortgage industry, ultimately serving the interests of banks, financial institutions, and their clientele.

OBJECTIVE OF THE PROJECT

The main objective of our project is to develop and evaluate the home origination model using advanced analytical tolls and banking data. In terms of analytical tools in this project we are using SAS Viya. We also aim to segment potential home loan borrowers based on their financial profiles, credit scores and risk factors. Similarly, we aim to analyse and understand the home loan climate in Australia, identifying trends, challenges, and opportunities. Also, we want to investigate the impact of regulatory and economic factors on home loan origination and affordability. Enhancing team's skills in utilizing the tools and understanding the concept of managing this kind of project is also one of the major objectives of this project.

RESTATEMENT OF THE SCOPE AND REQUIREMENTS

During the course of this project execution, there were revisions made to the initial plan regarding the prediction of delinquency for each data point in SAS Viya. Originally, the project scope included performing delinquency prediction at a granular level for individual data points within the dataset. However, as the project progressed, it became evident that this approach was exceedingly time-consuming and resource-intensive, posing challenges in terms of feasibility and practicality within the project timeline.

After careful consideration and consultation with the sponsors, it was decided to revise the approach and streamline the delinquency prediction process. Instead of predicting delinquency for each data point individually, the focus shifted to performing prediction for the entire dataset. This modification allowed for a more efficient and manageable execution of the prediction task, aligning with project objectives and constraints. Also, the separate dashboard to view the single point data was also decided not to pursue as it will be time consuming and was not in scope range of the project.

The decision to revise the approach was documented and communicated to all relevant stakeholders, including project sponsors, team members, and mentor. Any necessary approvals were obtained to proceed with the revised plan, ensuring alignment with project goals and objectives.

LITERATURE REVIEW

In 1993, Australians borrowed around $100bn or 22% of GDP in housing finance from lenders. Ten years later, housing borrowings expanded to $385bn.[3] Whereas currently according to the ABS lending indicators, the value of loan commitments in Australia during the June 2023 quarter alone was $24.6 billion dollars.[4]

A loan is not unsuitable if it meets the consumer's requirement and objectives, and the consumer has the capacity to repay the loan without experiencing substantial hardship. [3] As home ownership is an important goal for many Australians, many Australians consider it important to retain their own homes and, therefore, meet their obligations under housing loans, even if they are experiencing financial stress. More than 66% of Australian household live in owner-occupied dwellings. Of these homeowners, 44% own their properties outright (29.5% of all households), without a mortgage loan. The proportion with of homeowners without a mortgage has fallen in the past 10 years, while the proportion of households with a mortgage has risen, as has the proportion of renters as a percentage of total households. These changes, reflect the effects of prolonged period of strong property price growth, which has affected housing affordability, affecting Australian mortgage industry and lenders.[6]

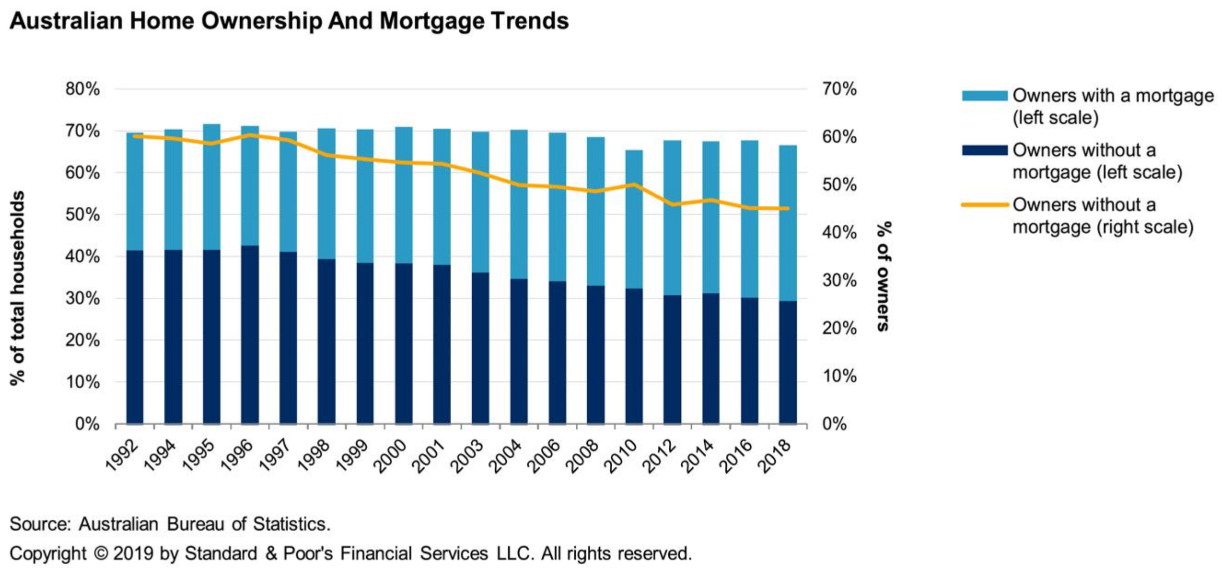

FIGURE 1 AUSTRALIAN HOME OWNERSHIP AND MORTGAGE TRENDS

The chart shows the percentage of homeowners with a mortgage in Australia has been increasing over time. This suggests that people are increasingly relying on mortgages to finance their home purchase. We can also see the decline in home ownership without a mortgage, indicating the problem such as rising house prices, making it more difficult for people to save up enough money for a down payment. Based on above chart, we can assume that there will be number of factors that will be affecting this industry like increased use of technology, more competition among lenders, and a growing focus on affordability.

In 90's we witnessed boom times in mortgage market fuelled by easy access to credit, deregulation, and a stable economy. Government incentives, such as tax breaks and grants, encouraged homeownership and property investment. Innovating loan products like low-document loan and interest only loans were introduced, leading to rapid expansion of credit. However, Tax regulations failed to keep pace with this credit expansion, resulting in risky lending practices. But after the financial crisis, market experienced some correction with credit tightening and house price growth slowing down. Borrowers and lenders became more cautious due to economic uncertainty, leading to stricter lending standards. Now, Australian mortgage market is focusing on stability rather than rapid expansion. [1]

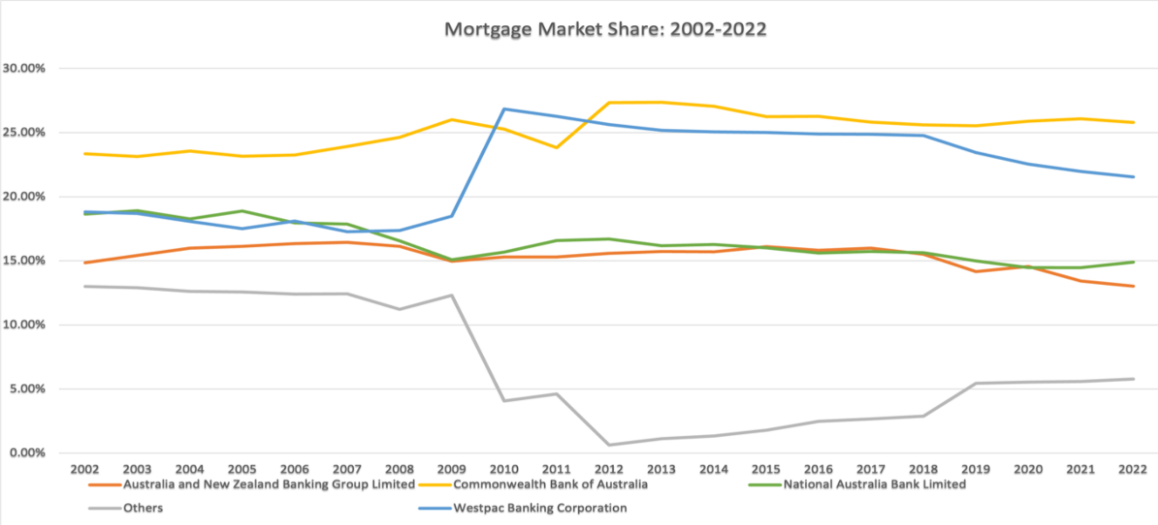

FIGURE 2 MORTGAGE MARKET SHARE 2002 - 2022

(Data Source: Australian Prudential Regulation Authority (APRA))

The above chart shows the market share of four major Australian banks: Commonwealth Bank of Australia, Australia and New Zealand Banking Group Limited, National Australia Bank Limited and Westpac Banking Corporation. We can see the commonwealth bank appears to have had the highest market share throughout the period, though it has fluctuated some. Westpac Banking Corporation seems to have had the second-highest market share overall, also with some variation over the years. The market share of National Australia Bank Limited and Australia and New Zealand Banking Group Limited appears to be lower and more stable than the other two banks.

In the Australian mortgage market, loan products cater to diverse borrower needs. Some popular options include:

1. Variable Rate: Interest rates fluctuate with the RBA cash rate, offering flexibility.
2. Honeymoon Mortgage: Offers discounted variable rates for a fixed period.
3. Fixed Rate: Provides repayment certainty for a fixed period.
4. Home Equity Loan: Allows access to funds using property equity as collateral.
5. Home Loan Packages: Bundled features at a discounted rate for added convenience.
6. Other Options: Including construction loans, land loans, and investment property loans.

These options offer flexibility, stability, and convenience to borrowers based on their financial goals and preferences. [7]

Loan origination models and delinquency prediction is one of the major tasks that needs to be done by financial institutions before disbursement of loan. Various machine learning algorithms are commonly employed to analyse data and make predictions. Here are some of the key algorithms used in this domain:

1. ***Logistic Regression:***
    - Logistic regression is widely used for binary classification problems, such as predicting whether a loan applicant will default or not.
    - It estimates the probability of a binary outcome based on one or more independent variables.
    - Logistic regression provides interpretable results and is relatively simple to implement.
2. ***Decision Trees:***
    - Decision trees are versatile and intuitive models used for both classification and regression tasks.
    - In the context of loan origination, decision trees can be used to segment borrowers based on various attributes such as income, credit history, and loan amount.
    - Decision trees are easy to understand and visualize, making them valuable for interpreting model decisions.
3. ***Random Forest:***
    - Random Forest is an ensemble learning technique that builds multiple decision trees and combines their predictions.
    - It improves prediction accuracy and reduces overfitting compared to individual decision trees.
    - Random Forest can handle large datasets with high-dimensional feature spaces effectively.
4. ***Gradient Boosting Machines (GBM):***
    - GBM is another ensemble learning method that builds trees sequentially, with each tree correcting errors made by the previous one.
    - It's known for its high predictive accuracy and robustness against overfitting.
    - GBM models are often used in loan origination for their ability to capture complex relationships between borrower characteristics and loan delinquency.
5. ***Support Vector Machines (SVM):***
    - SVM is a supervised learning algorithm used for classification and regression tasks.
    - It works well for both linearly separable and non-linearly separable datasets.
    - SVMs are particularly useful when dealing with high-dimensional feature spaces and can handle complex decision boundaries.
6. ***Neural Networks:***
    - Neural networks, especially deep learning models, have gained popularity in recent years for their ability to learn complex patterns from data.
    - In loan origination, neural networks can be used to analyse large amounts of borrower data and predict the likelihood of default.

- However, they often require large amounts of data and computational resources for training.

These algorithms can be combined or used individually depending on the specific requirements of the loan origination model and delinquency prediction task. Additionally, feature engineering, data pre-processing, and model evaluation techniques play crucial roles in building accurate and robust predictive models in this domain [8].

METHODOLOGY

In this project, we employed a combination of literature review and empirical analysis to achieve our objectives. Our research methodology involved reviewing academic papers and industry reports to gain insights into previous models used to predict delinquency and trends in the Australian mortgage market. Additionally, we utilised empirical data analysis techniques to analyse the HMEQ dataset and derive meaningful conclusions.

The primary source of data for our analysis was the HMEQ dataset, which was on SAS platform. This dataset contains information on recent applicants granted credit for home equity lines of credit. We employed data pre-processing techniques to clean and prepare the dataset for analysis. Subsequently, we utilised predictive modelling tools to build credit scoring models based on the collected data. Our analysis also involved statistical techniques to assess model performance and interpret results.

In addition to the tools and technologies mentioned, we adopted the Agile methodology as our project management framework.

Here's how we applied Agile principles in our project:

1. ***Iterative Development:*** We divided the project into multiple iterations, known as sprints, each lasting for a fixed duration. During each sprint, we focused on completing specific tasks and delivering incremental value to stakeholders.
2. ***Adaptive Planning:*** Instead of creating a detailed project plan upfront, we embraced adaptive planning, allowing us to adjust our approach based on changing requirements, feedback, and priorities. This enabled us to respond quickly to evolving project needs and deliverables.
3. ***Cross-Functional Collaboration:*** Our project team comprised individuals with diverse skills, fostering cross-functional collaboration. We held regular meetings, known as stand-ups or scrum meetings, to discuss progress, address challenges, and ensure alignment among team members.
4. ***Stakeholder Involvement:*** Throughout the project, we maintained open communication channels with the project sponsor and project mentor. We solicited feedback at regular

intervals, allowing us to validate assumptions, prioritize features, and ensure that the project outcomes aligned with stakeholder expectations.

5. ***Continuous Improvement:*** We prioritized continuous improvement by conducting regular retrospectives at the end of each sprint. During these retrospectives, we reflected on our performance, identified areas for improvement, and implemented corrective actions to enhance our processes and practices.

DATA ANALYSIS

DATASET INFORMATION

The Home Equity dataset (HMEQ) contains baseline and loan performance information for 5,960 recent home equity loans.

The target (BAD) is a binary variable indicating whether an applicant eventually defaulted or was seriously delinquent.

This adverse outcome occurred in 1,189 cases (20%). For each applicant, 12 input variables were recorded. [9]

1. BAD: 1 = applicant defaulted on loan or seriously delinquent; 0= applicant paid loan (Target Variable)
2. LOAN: Amount of the loan request
3. MORTDUE: Amount due on existing mortgage (refers to a loan used to purchase real estate)
4. VALUE: Value of current property
5. REASON: DebtCon = debt consolidation (involves combining multiple debts into a single loan or payment to simplify management and potentially lower interest rates) .Home Imp = home improvement (involves enhancing homes through renovations and upgrades to improve their appearance, functionality, and value)
6. JOB: Occupational categories
7. YOJ: Years at present job
8. DEROG: Number of major derogatory reports (count of credit accounts where the borrower has failed to make payments on time)
9. DELINQ: Number of delinquent credit lines (the count of significant negative credit reports)
10. CLAGE: Age of oldest credit line in months (since the establishment of the borrower's earliest credit account or trade line)
11. NINQ: Number of recent credit inquiries (count of new credit accounts or trade lines established by the borrower in a specified timeframe)
12. CLNO: Number of credit lines (the count of various credit accounts or trade lines associated with the borrower's credit history)
13. DEBTINC: Debt-to-income ratio (the comparison between a borrower's monthly debt payments and their gross monthly income)

PRESENTATION OF ANALYTICAL FINDINGS

Column	Data Type	Null	Non-null Count	Mean	Median	Mode	Standard Devia...
BAD	double		5,960	0.20	0.00	0.00	0.40
CLAGE	double	5.17% (308)	5,652	179.77	173.47		85.81
CLNO	double	3.72% (222)	5,738	21.30	20.00	16.00	10.14
DEBTINC	double	21.26% (1,267)	4,693	33.78	34.82		8.60
DELINQ	double	9.73% (580)	5,380	0.45	0.00	0.00	1.13
DEROG	double	11.88% (708)	5,252	0.25	0.00	0.00	0.85
JOB	varchar	4.68% (279)	5,681			Other	
LOAN	double		5,960	18,607.97	16,300.00	15,000.00	11,207.48
MORTDUE	double	8.69% (518)	5,442	73,760.82	65,019.00	42,000.00	44,457.61
NINQ	double	8.56% (510)	5,450	1.19	1.00	0.00	1.73
REASON	varchar	4.23% (252)	5,708			DebtCon	
VALUE	double	1.88% (112)	5,848	101,776.05	89,235.50	60,000.00	57,385.78
YOJ	double	8.64% (515)	5,445	8.92	7.00	0.00	7.57

FIGURE 3 VARIABLES IN THE DATASET

This table includes the different variables such as CLNO, LOAN, REASON, CLAGE and so on. Similarly, the table includes datatype for each variable, Percentage of Null value, Mean, Median, Mode and Standard Deviation. This table main purpose is to indicate the percentage of Null values in each variable. We can see 21.26% of Debt-to-income ratio data has null value which significantly impact the outcome of the model. Similarly, we have DEROG, DELINQ variables which have higher value of missing value percentage.

There are different algorithms which can be used to train the models. And SAS Viya platform provides us with wide range of options while building the model.

The 'Model Studio – Build Model' is the interface where we can initiate the model building process. Here we can create multiple models with different parameters and compare each other performance.

One of our targets in this project was to create 3 different models that would give us the results on the basis of their own trained criteria.

Filter

Variable Name	Label	Type	Role	Level	Order
BAD		Numeric	Target	Binary	Default
CLAGE		Numeric	Input	Interval	Default
CLNO		Numeric	Input	Interval	Default
DEBTINC		Numeric	Input	Interval	Default
DELINQ		Numeric	Input	Nominal	Default
DEROG		Numeric	Input	Nominal	Default
JOB		Character	Rejected	Nominal	Default
LOAN		Numeric	Input	Interval	Default
MORTDUE		Numeric	Input	Interval	Default
NINQ		Numeric	Input	Nominal	Default
REASON		Character	Rejected	Binary	Default
VALUE		Numeric	Input	Interval	Default
YOJ		Numeric	Input	Interval	Default

Multiple Variables
Role: Rejected
Level: Mixed values
Order:
Transform:
Impute:

FIGURE 4 VARIABLE SELECTION FOR MODEL TRAINING

To achieve this target, we have rejected the variables that we have not been and using the remaining ones to train the model. To reject the variables, we can select the variables and choose the 'Role' to rejected.

MODEL SEGMENTATION

One model will be trained using all the variables (except Job and Reason as they do not play any effective role for the outcome).

The other model (Model 2) will be trained using the variables:

- CLAGE: Age of oldest trade line in months
- CLNO: Number of credit lines
- DEROG: Number of major derogatory reports
- DELINQ: Number of delinquent credit lines
- NINQ: Number of recent credit lines

The last model (Model 3) will be trained using the variables:

- DEBTINC: Debt to income ratio
- LOAN: Amount of requested loan
- MORTDUE: Amount due on existing mortgage.
- VALUE: Valuation of current property
- YOJ: Years in current job

Regarding segmentation, while the variables used in Models 2 and Model 3 may seem similar in nature, they likely represent distinct dimensions of borrower characteristics and financial situations. Model 2 focuses more on credit history and behaviour, such as the number of credit lines, derogatory reports, and delinquent credit lines. Model 3 emphasizes financial status and property valuation, including debt-to-income ratio, loan amount, existing mortgage amount, property value, and years in the current job. By training models with different sets of variables, the segmentation may arise naturally from the inherent patterns present in the data. Each model may identify different clusters or segments of borrowers based on their unique combinations of features. These segments could represent groups of borrowers with similar risk profiles, financial situations, or borrowing behaviours.

In the initial stage, we used the template named as "Advance template for class target" which is provided by SAS. This pipeline uses different nodes like Neural network, forest, gradient boosting, and ensemble model.

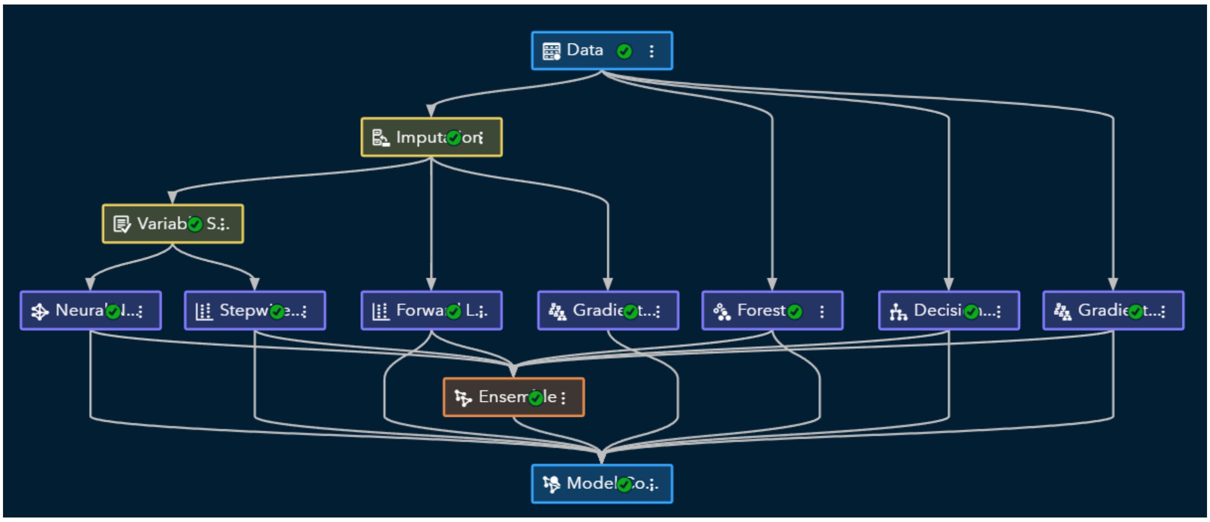

FIGURE 5 ADVANCE TEMPLATE FOR CLASS TARGET

Through this first training, we found out that our dataset is not suitable for feeding into the Neural Network models. One of the reasons is the size of the data. In order to leverage the proper use of the Neural Network architecture, a huge amount of dataset is required which is not with us at the current moment.

So, we came to the decision to use the statistical machine learning algorithms and develop a custom pipeline for training our model.

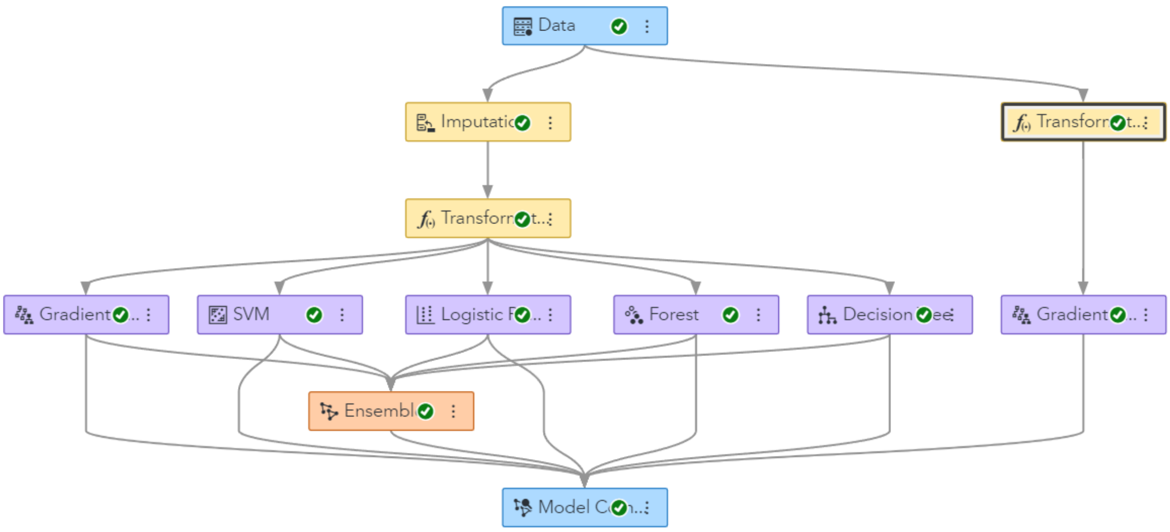

FIGURE 6 CUSTOM TEMPLATE FOR CLASS TARGET

In the next step we started experimenting with new features that are available in the SAS Viya platform. There are few pre-processing steps that can be performed before feeding the data to the algorithm. So, we added the nodes like:

- ***Imputation:*** Since we have some columns/variables with missing values, it could affect the learning process of the model. So, we have the imputation node to impute those missing values.

- ***Transformation:*** This is also one of the important data processing steps which helps with problems like outliers and skewed data.
- ***Feature extraction:*** This node works on identifying and extracting the features which are important and have more impact to the target variable our output.

After the completing of pre-processing, we have fed those data into different algorithms like Decision Tree, Gradient Boosting, Logistic Regression, SVM, and Forest (Multiple Decision Tree). Each of these algorithms has their own patten of training.
The drag and drop option provided by SAS Viya was helpful in training the model using all these algorithms quickly and efficiently.

COMPARISON BETWEEN MODELS DEVELOPED

After creating multiple pipelines of multiple algorithms and training different models, we have found that Gradient Boosting is the most suitable approach for our dataset. And the required pre-processing technique is 'Transformation' with 'Inverse Square Root' technique used for transforming the variables.

Champion	Name	Algorithm Name	KS (Youden)	Misclassification Rate
	Gradient Boosting (1)	Gradient Boosting	0.7753	0.0822
	Gradient Boosting	Gradient Boosting	0.5384	0.1326
	Forest	Forest	0.5362	0.1493
	Ensemble	Ensemble	0.5173	0.1661
	Logistic Regression	Logistic Regression	0.3280	0.1862
	Decision Tree	Decision Tree	0.2648	0.1695
	SVM	SVM	0.1934	0.1862

FIGURE 7 MODEL COMPARISION

The different trained pipelines can be viewed in SAS Viya through the 'Pipeline Comparison' tab. It also shows us the champion algorithm.
We can also have a graphical representation of the model's comparison. We can plot various types of figures like:

- Lift:
- ROC
- Accuracy
- FIT statistics

In the lift information table, taking the percentile analysis of the probability of the events, the number of events is allocated to each bin. Below, we have visualized the 'Response Percentage' which is the number of events divided by the number of observations in each bin. We can also evaluate the model performance on another context like:

- Captured Response Percentage
- Cumulative Captured Response Percentage
- Cumulative Response Percentage

- Cumulative Lit
- Gain
- Lift

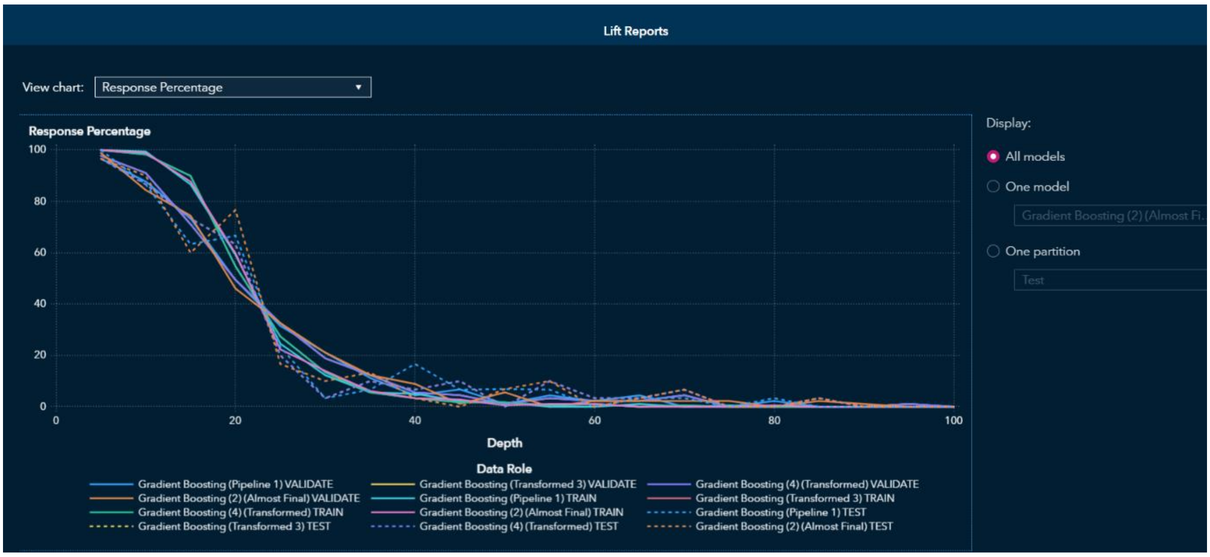

FIGURE 8 LIFT CURVE

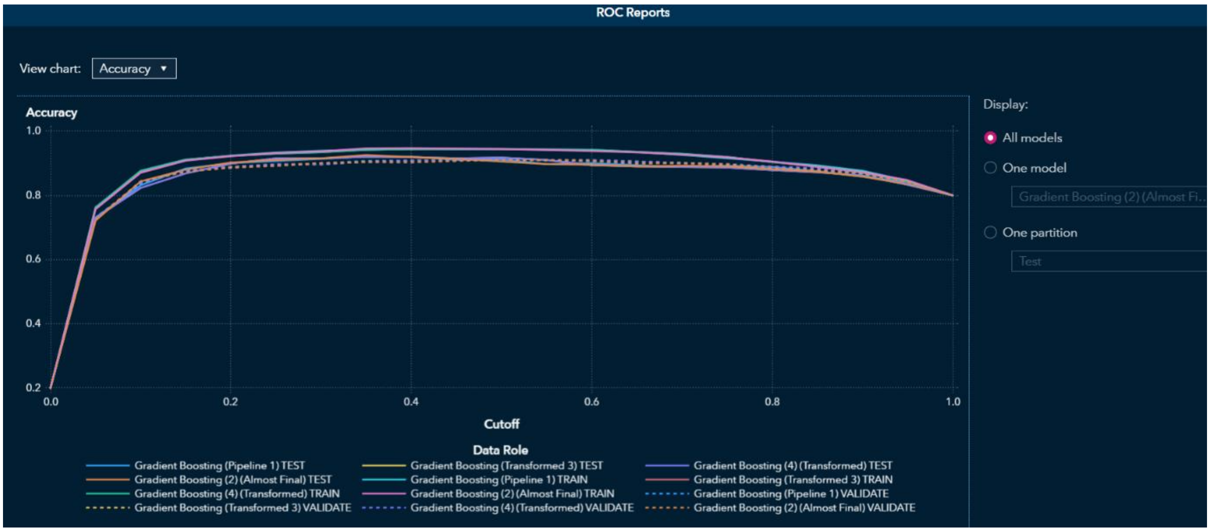

FIGURE 9 ROC CURVE

In the accuracy plot, we can visualize the accuracy level of all the models during the training process. We can even check on the report on basis of individual model by clicking the ‘One model’ option and choosing the desired model. With the same approach, we can visualize the accuracy score on the training dataset, validation dataset or testing dataset.

The ‘Fit Statistics’ table gives us with various statistical information of all the models.

Statistics ...	Train: Gr...	Validate:...	Test: Gra...	Train: Gr...	Validate:...	Test: Gra...	Train: Gr...	Validate:...	Test: Gra...	Train: Gr...	Validate:...	Test: Gra...
Area Under ROC	0.9773	0.9373	0.9235	0.9768	0.9410	0.9253	0.9768	0.9410	0.9253	0.9756	0.9387	0.9238
Average Squared Error	0.0423	0.0670	0.0667	0.0422	0.0645	0.0666	0.0422	0.0645	0.0666	0.0422	0.0662	0.0668
Divisor for ASE	3,576	1,788	596	3,576	1,788	596	3,576	1,788	596	3,576	1,788	596
Formatted Partition	1	0	2	1	0	2	1	0	2	1	0	2
Gamma	0.9677	0.9107	0.9005	0.9676	0.9162	0.9015	0.9676	0.9162	0.9015	0.9665	0.9130	0.8997
Gini Coefficient	0.9547	0.8745	0.8469	0.9535	0.8820	0.8507	0.9535	0.8820	0.8507	0.9512	0.8773	0.8476
KS (Youden)	0.8498	0.7538	0.7691	0.8489	0.7622	0.7753	0.8489	0.7622	0.7753	0.8512	0.7671	0.7628
KS Cutoff	0.2500	0.1500	0.2500	0.1500	0.1500	0.2500	0.1500	0.1500	0.2500	0.2500	0.1500	0.2000
Misclassification at Cutoff	0.0554	0.0889	0.0940	0.0545	0.0878	0.0822	0.0545	0.0878	0.0822	0.0551	0.0900	0.0872

FIGURE 10 FIT STATISTICS TABLE

Below we have identified some other visualization reports which can be achieved by the help of SAS Viya.

Variable Importance

Variable Label	Role	Variable Name	Training Importance	Importance Standard Deviati...	Relative Importance
Transformed DEBTINC	INPUT	INVSQRT_DEBTINC	17.9009	36.4089	1
	INPUT	DELINQ	4.3709	4.3253	0.2442
Transformed CLAGE	INPUT	INVSQRT_CLAGE	4.0351	3.0607	0.2254
Transformed VALUE	INPUT	INVSQRT_VALUE	3.8591	1.9396	0.2156
	INPUT	NINQ	3.0166	1.4162	0.1685
Transformed YOJ	INPUT	INVSQRT_YOJ	2.9598	1.3867	0.1653
	INPUT	DEROG	2.6918	2.4950	0.1504
Transformed CLNO	INPUT	INVSQRT_CLNO	2.3899	1.1640	0.1335
Transformed LOAN	INPUT	INVSQRT_LOAN	2.1171	1.1534	0.1183
Transformed MORTDUE	INPUT	INVSQRT_MORTDUE	1.9796	1.1103	0.1106

FIGURE 11 TRANSFORMED VARIABLE IMPORTANCE

The above figure gives us the information about which variables has been playing the important roles while training the model.

FIGURE 12 EVENT CLASSIFICATION REPORT

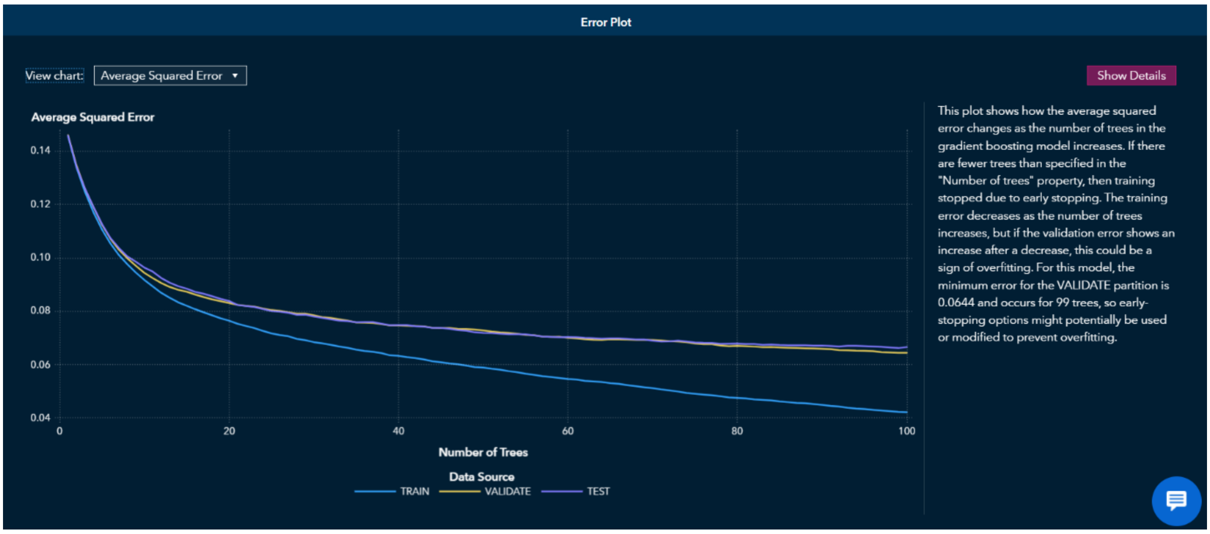

FIGURE 13 ERROR PLOT (AVERAGE SQUARED ERROR)

Case Studies and Applications

Case Study 1: Determinants of Mortgage Loan Delinquency: Application of Interpretable Machine learning

This paper used machine learning to determine factors influencing mortgage loan delinquencies. This study shows that credit score is one of the critical factors in mortgage lending decisions. Their focus is on 90-day delinquency within the first 12 months after loan origination. In this study they have applied a modified Gradient Boosting approach (XGBoost) to data from 2013Q1 to 2017Q4. First, they found that XGBoost's prediction accuracy is higher than Logistic regression. Second, they used Permutation Feature Importance method to find the most important factors impacting delinquency. Finally, they used Interpretable Machine Learning to quantify the effect of each factor on delinquency probability. Borrower's credit score, Federal Funds Rate, and original interest rate are found as the most important features in the model. Lenders access borrower's creditworthiness using credit scores, which are influenced by factors such as payment history, outstanding debts and credit utilization. Lower credit scores often lead to stricter loan terms and high interest rates, increasing the risk of delinquency. Similarly, the interest rate at which the mortgage is originated plays a significant role in determining monthly payments and affordability for borrowers. Higher interest rates can strain borrowers' finances, increasing the likelihood of delinquency, especially in times of economic stress. Debt-to-Income Ratio (DTI) measures the proportion of a borrower's monthly income that goes toward debt payments, including the mortgage. High DTI ratios indicate higher financial strain, which can increase the risk of delinquency, particularly if borrowers experience income disruptions or unexpected expenses. Original Loan-to-Value (OLTV) and Original Combined Loan-to-Value (OCLTV): These ratios represent the percentage of the property's value financed by the loan at origination. Higher OLTV and OCLTV ratios indicate higher leverage and potentially greater risk of default, especially if property values decline or borrowers experience financial difficulties. Loan Modifications to mortgage terms, such as extending the loan term or reducing interest rates, can impact delinquency risk. While modifications aim to assist struggling borrowers, they may also signal financial distress and increase the likelihood of future default. [10]

CASE STUDY 2: AN EMPIRICAL COMPARISON OF CLASSIFICATION ALGORITHMS FOR MORTGAGE DEFAULT PREDICTION: EVIDENCE FROM A DISTRESSED MORTGAGE MARKET

In this study, various modelling approaches were evaluated for predicting future mortgage default status, with boosted regression trees showing significant outperformance compared to logistic regression. This suggests that boosted regression trees can be a valuable tool for assessing mortgage credit risk. The main findings of the study regarding the performance of different classification algorithms for mortgage default prediction are as follows:

1. Boosted regression trees significantly outperformed logistic regression in predicting future mortgage default status.

2. The selected modelling approaches, including boosted regression trees, random forests, penalised linear, and semi-parametric logistic regression models, exhibited varying degrees of predictive power.

3. The study suggests that boosted regression trees can be a valuable addition to the current toolkit for assessing mortgage credit risk by banks and regulators.

These findings highlight the potential benefits of utilizing advanced modelling techniques, such as boosted regression trees, in improving the accuracy of mortgage default prediction models compared to traditional methods like logistic regression. The analysis in the study included four portfolios of over 300,000 Irish owner-occupier mortgages. Therefore, a total of four portfolios were evaluated, each consisting of over 300,000 mortgages.

The results of this study can be applied by banks and regulators in assessing mortgage credit risk and making informed decisions in the following ways:

1. Improved Predictive Models: Banks can enhance their credit risk assessment models by incorporating advanced techniques like boosted regression trees, which have shown superior performance in predicting mortgage default status compared to traditional methods like logistic regression.

2. Risk Management: By utilizing more accurate predictive models, financial institutions can better identify high-risk mortgage loans and take proactive measures to mitigate potential defaults. This can help in reducing overall credit risk exposure and improving portfolio quality.

3. Regulatory Compliance: Regulators can benefit from the study's findings by encouraging banks to adopt more sophisticated modelling approaches for mortgage credit risk assessment. This can lead to more robust risk management practices and better compliance with regulatory requirements.

4. Capital Adequacy: Accurate assessment of mortgage credit risk is crucial for determining regulatory capital requirements. By using advanced classification algorithms, banks can more precisely estimate the probability of default, leading to more appropriate setting of capital buffers to cover potential losses.

5. Loan Pricing Strategy: The outputs of improved credit scoring models can also contribute to the implementation of risk-adjusted loan pricing systems. Banks can adjust their pricing strategies based on the predicted creditworthiness of borrowers, leading to more tailored and competitive loan offerings.

Overall, the study's findings provide valuable insights for banks and regulators to enhance their mortgage credit risk assessment processes, leading to more informed decision-making, better risk management, and improved regulatory compliance. [11]

CASE STUDY 3: MODELLING AND FORECASTING MORTGAGE DELINQUENCY AND FORECLOSURE IN THE UK

In this paper, Janine Aron and John Muellbauer discusses the modelling and forecasting of mortgage delinquency and foreclosure in the UK. They present new aggregate models based on quarterly data from 1983 to 2014, focusing on factors such as loan quality, access to refinancing, and lenders' forbearance policies. The key factors influencing mortgage delinquency and foreclosure rates in the UK, as discussed in the study by Janine Aron and John Muellbauer, include:

1. Loan Quality: The quality of the loans, including factors such as loan-to-value ratio, interest rates, mortgage terms, and borrower characteristics, plays a significant role in determining the likelihood of delinquency and foreclosure.
2. Access to Credit: The availability of credit and access to refinancing options can impact borrowers' ability to meet their mortgage payments and avoid delinquency.
3. Lenders' Forbearance Policies: Lenders' policies regarding forbearance, which involves temporary relief or suspension of mortgage payments, can influence the rates of delinquency and foreclosure by providing borrowers with assistance during financial difficulties.
4. Government Support: Government initiatives and income support programs aimed at assisting individuals facing mortgage payment difficulties can help mitigate delinquency and foreclosure rates.

5. Interest Rates: Fluctuations in interest rates, especially in an economy where most mortgages are at floating rates, can have a significant impact on borrowers' ability to make mortgage payments and may lead to higher delinquency and foreclosure rates.
6. Economic Conditions: Overall economic conditions, including factors like unemployment rates, income levels, and housing market trends, can also influence mortgage delinquency and foreclosure rates in the UK.

By considering these key factors and their interactions, researchers can better understand and forecast mortgage delinquency and foreclosure trends in the UK housing market. The findings of the study by Janine Aron and John Muellbauer on mortgage delinquency and foreclosure in the UK have several implications for credit risk stress testing and mortgage market regulation in the country:

1. Enhanced Risk Assessment: The study highlights the importance of considering factors such as loan quality, access to credit, and economic conditions in credit risk stress testing models. Regulators and financial institutions can use these insights to improve their risk assessment processes and better anticipate potential challenges in the mortgage market.

2. Policy Adjustments: The research underscores the need for regulators to adapt mortgage market regulations in response to changing economic conditions and lending practices. By understanding the dynamics of delinquency and foreclosure rates, policymakers can implement targeted measures to mitigate risks and protect borrowers and lenders.

3. Forbearance Policies: The study emphasizes the role of lenders' forbearance policies in influencing mortgage delinquency and foreclosure outcomes. Regulators may consider guidelines or standards for lenders to ensure fair and effective forbearance practices that support borrowers in financial distress while maintaining the stability of the mortgage market.

4. Government Intervention: The findings suggest that government support programs can have a significant impact on reducing delinquency and foreclosure rates. Policymakers may use this information to design and implement effective intervention strategies during economic downturns or periods of heightened financial stress.

5. Monitoring and Surveillance: The study underscores the importance of ongoing monitoring and surveillance of mortgage market trends and risk factors. Regulators can use the insights from this research to develop early warning systems and proactive measures to address emerging risks in the mortgage sector.

Overall, the findings of this study provide valuable insights for policymakers, regulators, and financial institutions in the UK to strengthen credit risk management practices, enhance regulatory frameworks, and promote stability in the mortgage market[12].

RECOMMENDATIONS

The inaugural Credit Stress Barometer for March 2023 shown clear signs of deteriorating credit Default Risk amongst Australian Consumers. We saw credit default risk of Australian consumer rise over 8% in the year from Jan 2023 - March 2023. Direct cause of this increased credit stress includes rising overdue repayments, higher rental obligations. These trends indicate that there is a need to be more cautious in-home loans given by banks. Similarly, for service provider for banks it is very important to consider too few aspects while making home loan origination model using technology like artificial Intelligence [13].

LOS (LOAN ORIGINATION SYSTEM): Loan origination system is defined as software that automates and streamlines the process of originating and processing loans, managing the entire loan lifecycle from application to disbursement and beyond.

Factors for Consideration:

- Ease of Use: The LOS should have a user-friendly interface for both lenders and borrowers, reducing time and effort and increasing loan application completion rates.
- Customization: The LOS should allow for customization to fit specific institution needs, including form fields, loan products, and workflow processes.
- Integration Capabilities: Seamless integration with other systems like core banking, CRM, and credit bureaus is crucial for efficiency, risk reduction, and informed decision-making.
- Scalability: The LOS should be able to handle high volumes of loan applications and adapt to the institution's growth without performance issues.
- Compliance: Compliance with regulations such as Know Your Customer (KYC) and Anti-Money Laundering (AML) is essential, along with security features to protect borrower information.

REPORT ON RESOURCES

During the execution of the project, a variety of resources were utilized to facilitate research, analysis, and implementation activities. These resources encompassed both digital tools and human capital, contributing to the successful completion of project milestones. Here's an overview of the resources used:

- Research Papers and Blogs:
  Various research papers and blogs related to home origination models, delinquency prediction, and mortgage market trends were accessed and analysed to gather insights and inform decision-making processes.
- SAS Viya:

SAS Viya, an advanced analytics platform, was employed for data pre-processing, model development, and predictive analytics tasks. Its robust features and capabilities facilitated efficient data analysis and modelling processes.

- Microsoft Excel:

  Microsoft Excel was utilized for data manipulation, visualization, and reporting purposes. It provided a versatile tool for organizing and analysing large datasets, generating insights, and presenting findings in a clear and concise manner.

- Microsoft Word:

  Microsoft Word served as the primary tool for documentation, including project reports, summaries, and presentations. It enabled the creation of professional-quality documents for internal and external communication.

- Human Resources:

  The project team comprised skilled professionals with expertise in data analysis, statistical modelling, and domain knowledge in the mortgage industry. Team members dedicated their time and expertise to research, analysis, model development, and collaboration efforts.

- Time Tracking Sheet:

  A time tracking sheet was used to record the hours spent by each team member on different project tasks. This allowed for accurate monitoring of resource allocation and helped in optimizing productivity and task prioritization.

Overall, the effective utilization of these resources, coupled with the dedication and expertise of the project team, contributed to the successful execution of the project and the achievement of its objectives.

REPORT ON OUTSTANDING ISSUES

Throughout the project duration, several challenges and unresolved issues were encountered, which may have implications for project completion and post-project operations. The key outstanding issues are as follows:

- Time Constraints:

  A major challenge faced during the project was time constraints. The project initiation could have been more prompt, allowing for better planning and allocation of resources. Additionally, the initial approach differed from the later stages, leading to adjustments and rework, which consumed valuable time.

- Scope Creep:

  One notable issue was scope creep, particularly concerning the attempt to predict delinquency for each data point individually. This task proved to be more time-consuming and resource-intensive than anticipated, diverting focus from core project objectives. Consequently, efforts had to be redirected, and the scope was adjusted to align with project constraints.

REPORT OF RISK MITIGATED

Throughout the project lifecycle, various risks were identified and effectively mitigated to minimize their potential impact on project outcomes.

Risk	Description	Mitigation Strategy
Scope Creep	Potential deviations from project objectives and timeline due to attempts to predict delinquency for each data point.	• Regular project reviews and stakeholder communication. • Evaluation of proposed scope changes against objectives and constraints. • Adjustment to ensure alignment and focus on priority tasks.
Time Constraints	Difficulty meeting strict deadlines and deliverables due to time limitations.	• Detailed project timeline with key milestones and deliverables. • Efficient resource allocation and task prioritization. • Regular progress reviews and adjustments to timelines.
Team Member Illness	Risk to project continuity and workload distribution due to team member illness or absences.	• Contingency plan for absences and disruptions. • Cross-training and knowledge sharing among team members. • Clear communication channels for seamless coordination and handover of responsibilities
Communication and Coordination Challenges	Potential for misunderstandings, delays, and inefficiencies due to communication and coordination issues.	• Regular team meetings, status updates, and progress reports. • Project management tools and platforms for streamlined

		communication and real-time updates
Technical Dependencies and Constraints	Limited access to SAS full version software and clean data hindering progress and impacting outcomes.	• Identification and implementation of alternative solutions and workarounds. • Collaboration with stakeholders and partners to address technical constraints and access resources. • Contingency plans for unexpected technical issues or dependencies
Imbalanced Dataset	Dataset with high percentage of missing value	• Missing values were imputed with mean value.

REPORT ON LESSON LEARNT

The project provided valuable insights and lessons learned that have contributed to our understanding of effective project management practices, technical implementation, and ethical considerations. The following highlights key lessons learned, emphasizing ethical considerations, successes, and areas for improvement:

INSIGHTS GAINED

- Ethical Considerations:
  One prominent ethical issue that emerged during the project was the importance of data privacy and confidentiality. As we worked with sensitive borrower information, ensuring the ethical handling and protection of data became paramount. This involved adhering to data protection regulations, obtaining necessary consent, and implementing robust security measures to safeguard against unauthorized access or misuse of data.
- Technical Implementation:
  The project underscored the significance of selecting appropriate tools and technologies to meet project objectives efficiently. Leveraging advanced analytics platforms such as SAS Viya enabled us to streamline data analysis and modelling processes, enhancing the accuracy and effectiveness of our predictive models. Additionally, the use of collaborative tools such as Microsoft team, OneDrive facilitated seamless communication and collaboration among team members, enhancing productivity and coordination.

- Project Management Practices:

  Effective project management practices played a crucial role in project success, including clear communication, regular progress monitoring, and proactive risk management. Maintaining open lines of communication ensured alignment and transparency among team members, while ongoing progress monitoring enabled timely identification of issues and adjustments to project plans as needed. Proactive risk management allowed us to anticipate and mitigate potential challenges, minimizing disruptions to project timelines and objectives.

REFLECTIONS

- Successes:

  One notable success of the project was the successful development and evaluation of sophisticated home loan origination models, which provided valuable insights into borrower behaviour and credit risk. Additionally, the collaborative efforts of the project team facilitated efficient problem-solving and knowledge sharing, contributing to project success.
- Areas for Improvement:

  Despite the overall success of the project, there were areas for improvement identified, including better alignment of project scope and objectives from the outset. Additionally, proactive identification and mitigation of risks could have been further strengthened to anticipate and address potential challenges more effectively. Furthermore, enhancing data governance and documentation practices would improve transparency and accountability in data management processes.

HANDOVER MATERIALS

As part of the project handover process, we are providing comprehensive documentation and materials to facilitate understanding, reference, and future use. These materials include:

- Project Report:

  A detailed project report encompassing case studies, past works related to the project, project findings, and conclusions. This report offers a comprehensive overview of the project's objectives, methodologies, and outcomes, serving as a valuable reference for stakeholders.
- Time Tracking and Task Allocation:

  Documentation outlining the allocation of hours for each task, including time tracking records and task breakdowns. This information provides insights into resource utilization, project progress, and task prioritization, facilitating future planning and optimization.

ID	Task Description	Duration (Hours)	Assigned To	Status	WEEK
0	Select Project and Submit Expression of Interest	100	Aakash, Birat, Prahlad, Shital	Completed	1
1	Develop Project Plan	50	Aakash, Birat, Prahlad, Shital	Completed	2 and 3
1.1	Define project scope and objectives	37	Aakash, Birat, Prahlad, Shital		
1.2	Identify project stakeholders	5	Aakash, Birat, Shital		
1.3	Create project schedule and start creating project backlog	10	Aakash, Prahlad		
1.4	Distribute roles among team members	1	Aakash, Birat, Prahlad, Shital		
1.5	Establish Communication Plan	2	Prahlad		
2	Research on the Homeloan and Mortgage Industry of Australia	150	Birat, Shital	Completed	2 to 5
3	Explore SAS Viya	300	Aakash, Birat, Prahlad, Shital	Concurrent Task	4 to 9
3.1	Locate and explore the HMEQ dataset	50	Aakash, Birat, Prahlad, Shital	Completed	
3.2	Prepare proposal with detailed plan	60	Aakash, Birat, Prahlad, Shital	Completed	
3.3	Do a thorough work breakdown and divide tasks of backlog into Sprints	10	Aakash, Prahlad	Completed	
4	Model Development	400	Aakash, Birat, Prahlad, Shital	Completed	7 to 10
4.1	Try various Machine Learning Models and select the best one		Aakash, Birat, Shital		
4.2	Evaluate model performance		Aakash, Birat		
4.3	Tune hyperparameters to improve model accuracy		Aakash, Birat, Prahlad, Shital		
4.4	Connect with SAS' REST API to integrate with custom web application	15	Prahlad	Closed	
4.5	Finalize Dashboard		Aakash, Prahlad		
5	Work on Poster, Presentation, and Final Report	250	Aakash, Birat, Prahlad, Shital	Ongoing	10 to 12
5.1	Poster		Birat, Shital		
5.2	Final Presentation		Aakash, Prahlad		
5.3	Final Report		Birat, Shital		

FIGURE 14 HOURS SPENT ON EACH TASK

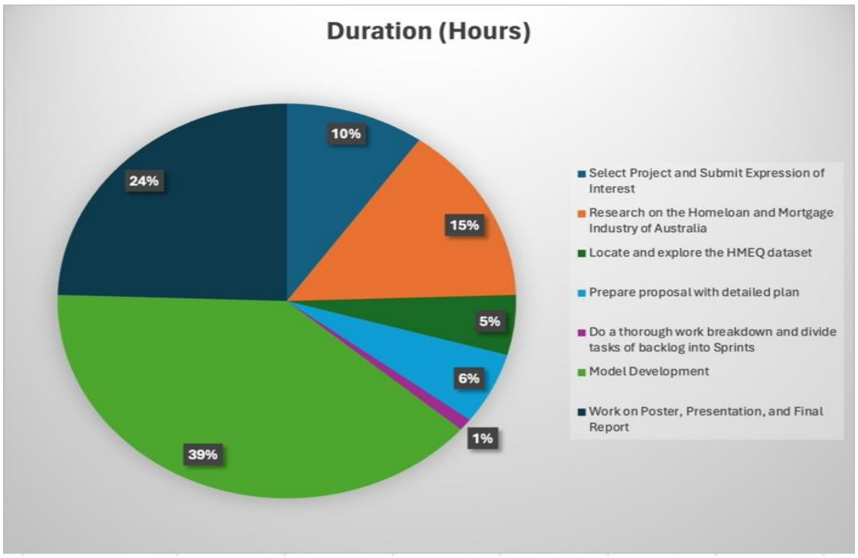

FIGURE 15 TIME DISTRIBUTION OF EACH TASK

- Backlogs and Action Items:

  A summary of backlogs and action items identified throughout the project lifecycle. This includes unresolved issues, pending tasks, and recommendations for further action, ensuring continuity and follow-up on outstanding matters.

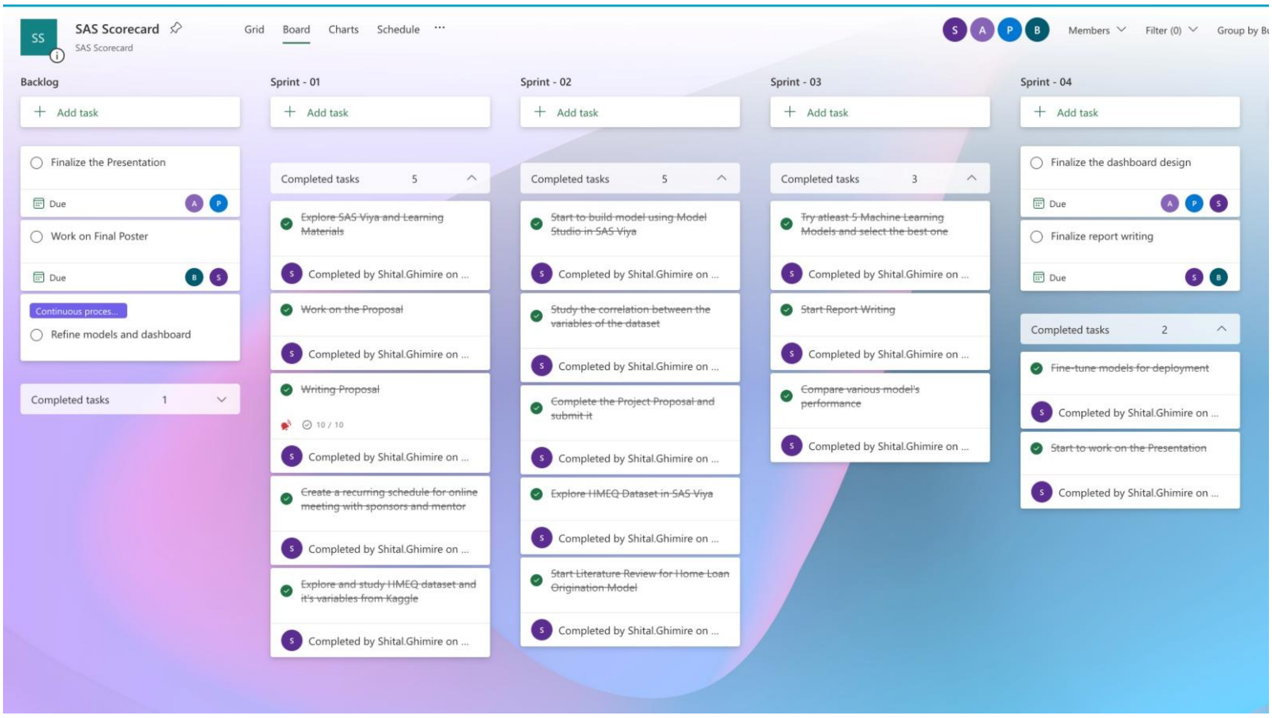

FIGURE 16 PROJECT BACKLOG

- Modelling Documentation:

  Detailed documentation of modelling methodologies, techniques, and algorithms employed during the project. This documentation serves as a reference for future modelling endeavours, providing insights into best practices, data pre-processing steps, and model evaluation techniques.

- Presentation Materials:

  Presentation slides prepared for the project presentation, including key findings, insights, and recommendations. These materials are designed to effectively communicate project outcomes and engage stakeholders during the presentation sessions.

- Supporting Materials and References:

  Supplementary materials, references, and resources used during the project, including research papers, articles, and data sources. These materials offer additional context and background information for stakeholders seeking further insight into specific aspects of the project.

PROJECT OUTCOMES

DELIVERABLES

In terms of project deliverables. The home loan origination model has been successfully developed, and final report consisting of literature review, market trends, case studies has been prepared and presentation slides has also been prepared. However, we encountered challenges in building the dashboard to visualize the scorecard. Integrating our model into the SAS Viya platform required a deeper understanding of the platform than initially anticipated. As a result, we made the decision to focus on predicting delinquency for the entire model rather than for individual data point.

ACHIEVEMENTS

- **Development of Segmented Models:** The successful development of predictive models utilizing major contributing variables marked a significant accomplishment for the team. These models facilitated the segmentation of potential home loan borrowers, allowing for more targeted and effective decision-making in the mortgage industry.

- **Understanding of Mortgage Market Dynamics:** Despite not having a finance background, the team achieved a comprehensive understanding of mortgage concepts and trends within the Australian mortgage market. This knowledge acquisition enabled informed analysis and decision-making throughout the project.

- **Diversified Learning in Project Management:** Through the project, the team acquired a diverse skill set in project management, including techniques such as example-based learning. This broader understanding of project management methodologies enhanced the team's ability to effectively execute tasks and achieve project milestones.

- **Networking Opportunities with Industry Professionals:** Engagement with sponsors from SAS provided valuable networking opportunities for the team. These interactions allowed for meaningful connections with industry professionals, fostering knowledge exchange and potential future collaborations.

- **Personal and Professional Growth:** The project served as a platform for the team's personal and professional growth. Learning experiences ranged from technical skill development in analytics to gaining insights into the complexities of the mortgage industry. Each achievement contributed to the team's overall development and readiness for future challenges in data analytics and banking.

ALIGNMENT WITH KPIS

S.N	Expected Outcome	KPIs	Measurement	Target	Alignment
1	Detailed Data Analysis Deport	Completeness of Data Analysis	Percentage of data elements analyzed comprehensively	100% data analysis presented in the report.	The final report includes a comprehensive analysis covering all the relevant elements, meeting the target of 100% completeness.
2	Robust Home Loan Origination Model	Model Accuracy	Percentage accuracy of the home loan origination model	Achieve a model accuracy rate of 90%	The model has achieved the maximum accuracy of 91%
3	Clear segmentation and tailored loan	Segmentation effectiveness	How well tailored loan product aligns	Customer satisfaction indicating alignment	The customer has been segmented into 2 parts. One based on their credit history and the other on their current financial strength.
4	Identification of key trends, challenges, and market opportunities	Trends and insights analysis	Number of key trends, challenges and opportunities identified	Adopting key trends and opportunities in the model and	The project identifies and addresses several key trends, challenges,

				mitigating challenges	and opportunities within the Australian mortgage market.
5	Visual engaging presentation	Audience engagement	Observation of audience engagement	Positive feedback and active participation from audience	Positive feedback has been received from the mentors and sponsors.
6	Demonstration of Enhanced skills	Skills proficiency	Evaluation of team member improvement in SAS tools	Demonstrate at least 25% improvement in skill proficiency.	All the team members are successfully able to get familiar with the SAS Viya and the Agile methodology.

QUALITY ASSURANCE MEASURES

Throughout the project lifecycle, rigorous quality assurance measures were implemented to ensure the accuracy, reliability, and integrity of the project outcomes. These measures adhered to the criteria and measures outlined in the Project Proposal and Plan.

- Data Validation: Thorough validation of data sources, ensuring data accuracy and consistency. This was very tough as the data was incomplete.
- Model Validation: Rigorous testing and validation of the developed models using robust methodologies such as cross-validation and holdout validation to ensure their reliability and generalizability.
- Peer Review: Regular peer review sessions were conducted to evaluate and validate the analysis methods, findings, and recommendations, enhancing the credibility and robustness of the project outcomes.

- Documentation: Comprehensive documentation of all processes, methodologies, and findings, ensuring transparency and reproducibility of the project results were carried out.
- Feedback Mechanism: Implementation of a feedback mechanism to solicit feedback from mentor and project sponsors, allowing for continuous improvement and refinement of project deliverables.

IMPACT

This project has had a significant impact on its intended audience, stakeholders, and the organization it serves. Some of the key impacts include:

- Improved Decision-Making: The development of robust home loan origination models and segmentation strategies can help banks and financial institutions to make more informed and data-driven decisions in approving loans and managing risks.
- Enhanced Risk Management: The implementation of predictive models and segmentation techniques will strengthen risk management practices within the mortgage industry, enabling lenders to identify and mitigate risks more effectively.
- Enhanced Customer Experience: Tailored loan products and services based on clear segmentation will result in a more personalized and satisfying experience for borrowers, improving customer satisfaction and loyalty.
- Strategic Insights: The analysis of key trends, challenges, and market opportunities has provided strategic insights to industry stakeholders, enabling them to adapt their strategies and offerings to meet evolving market demands.
- Professional Development: This project has also contributed to the professional development of team members, enhancing our skills in data analytics, project management, and stakeholder engagement, thus preparing us for future challenges in the field.

Overall, the project has made a positive impact on the mortgage industry, fostering innovation, improving decision-making processes, and driving positive outcomes for all people involved.

CONCLUSION

In conclusion, the project "Developing a Data-Driven Home Loan Origination Scorecard: Enhancing Risk Management and Decision-Making in the Australian Mortgage Industry" has been an amazing journey for us which is aimed at minimising risk and decision-making processes within the Australian mortgage sector. With the use of data analytical tool SAS Viya, the project has achieved significant milestones in these 13 weeks journey.

The development of predictive home loan origination models, coupled with clear segmentation strategies, marks a significant achievement in the project's journey. Additionally, the identification of key trends, challenges, and market opportunities has equipped stakeholders with strategic insights to navigate the dynamic landscape of the Australian mortgage market effectively.

Despite challenges encountered, such as the complexity of building a visualisation dashboard and the need for continuous learning in SAS Viya integration, the project team exhibited adaptability and resilience, ultimately delivering impactful outcomes.

Looking ahead, the insights gained from this project will serve as a foundation for future advancements and innovations in the field of data analytics and banking. As technology continues to evolve and new challenges emerge, the lessons learned, and achievements made in this project will continue to guide us throughout our career towards a more resilient and prosperous future.

BIBLIOGRAPHY

[1] M. B. Yanotti, “A review of the Australian Mortgage Market,” Tasmanian School of Business and Economics , 2014.

[2] J. S. Hubbard, “Responsible Lending: An international Landscape,” *Consumers International,* 2013.

[3] Macquaries Research Equities, “Australian Mortgage Market,” Let the Good times Roll, 2003.

[4] “Australian home loan statistics 2023,” 12 02 2024. [Online]. Available: https://ownhome.com/articles/australian-home-loan-statistics-2023#:~:text=The%20lending%20indicators%20show%20that,cost%20pressures%20of%20rate%20hikes.. [Accessed 05 04 2024].

[5] J. S. Hubbard and E. McNess, “The Australian Responsible Lending Act: The verdict is cautiously optimistic for the consumer,” *Consumer International.*

[6] E. Kitson, N. Coneybeare and A. E. Steenson, “An overview of Australia's Housing Market and Residential Mortgage-Backed Securities,” S&P Global Ratings, 2019.

[7] H. M. Karamujic, “A CLASSIFICATION OF HOME LOAN PRODUCTS IN AUSTRALIA,” in *Refereed proceeding of the 15th Annual Pacific Rim Real Estate Society (PRRES) Conference* , Sydney, Australia, 2009.

[8] S. S and J. L. E. P, “A COMPARATIVE STUDY ON MACHINE LEARNING ALGORITHMS FOR LOAN APPROVAL PREDICTION ANALYSIS,” *International Research Journal of Modernization in Engineering Technology and Science,* vol. 04, no. 12, pp. 565-569, 2022.

[9] “Home loan risk Prediction,” Kaggle, 17 08 2023. [Online]. Available: https://www.kaggle.com/code/harshitkumarsaxena1/home-loan-risk-prediction.

[10] T. Farzad, “Determinants of Mortgage Loan Delinquency: Application of Interpretable Machine Learning,” University of California, Riverside, California, 2013.

[11] T. Fitzpatrick and C. Mues, “An empirical comparison of classification algorithms for mortgage default prediction:evidence from a distressed mortgage market,” *European Journal of Operational Research,* vol. 01, pp. 1-13, 2015.

[12] J. Aron and J. Muellbauer, “Modelling and forecasting mortgage delinquency and foreclosure in the UK,” *Journal of Urban Economics,* pp. 32-53, 2016.

[13] M. Landgraf, “Australian's Credit Stress Barometer,” Empowering Intelligent Decision Making, 2023.