The Impact of Analytical Techniques in Credit Risk Management

Thesis submitted to the University of Nottingham for the degree of BSc (Hons) Finance, Management and Business Analytics

April, 2025

SIA TUNG YEE
20511127

Supervisor(s): Professor Seow Hsin-Vonn

Business (NUBS)
University of Nottingham

Word Count: 7923

Table of Contents

Abstract

This research examines the effectiveness of both traditional and modern techniques in evaluating credit risk, with a brief review of the evolution in credit risk assessment techniques and specifically selecting and focusing on logistic regression, decision trees, and random forests. While logistic regression is valued for its simplicity and interpretability, its ability to predict outcomes is limited in comparison to more advanced models. Decision trees, on the other hand, provide a structured, rule-based approach but are susceptible to overfitting, which reduces their accuracy. In contrast, random forests deliver superior prediction accuracy, although they come with the trade-off of lower interpretability.

The study further investigates the potential combination of traditional and modern analytical models to improve credit risk evaluation. By integrating these methods, financial institutions can benefit from a more balanced approach that enhances both prediction accuracy and transparency. This hybrid approach allows organisations to make more informed and transparent financial decisions, improving the overall management of credit risk.

Acknowledgements

First, I would like to sincerely express my gratitude to my supervisor, Dr. Seow Hsin-Vonn, for her continuous support and guidance throughout the process of writing this dissertation. Her expertise and thoughtful suggestions on this topic were crucial in keeping me on the right track and enhancing the quality of my research. I am especially grateful for her patience and the time she dedicated to weekly meetings and reviewing my progress along the way, offering valuable feedback that helped me better understand the topic that I have chosen.

Next, I would also like to express my appreciation to my fellow coursemates. Our shared encouragement and determination created an environment where we pushed each other to perform at our best. Last but not least, I am deeply grateful to my parents for their unwavering support both emotionally and financially. Their belief in the importance of education and their encouragement throughout my studies have been a big part in shaping who I am now and have made this journey possible.

Declaration of Authorship

I, Tung Yee Sia, declare that the dissertation entitled *"The Impact of Analytical Techniques in Credit Risk Management"* and the work presented in it are my own. I confirm that:

- this work was done wholly and mainly while in candidature for a research degree at this University;
- where any part of this dissertation has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this dissertation is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Signed: 

Date: 15th April 2025

List of Figures

List of Tables

Correction Sheet

1. Introduction

1.1 Research Background

Credit risk management has long been a crucial part of banking, responsible for assessing the probability that borrowers will meet their financial obligations. It plays a vital role in helping financial institutions manage profitability, minimise risk, and ensure long-term stability by maintaining sufficient liquidity, improving lending decisions, and complying with regulations such as the Basel Accords (BIS, 2011). By leveraging data-driven insights, financial institutions can make informed decisions, including loan approvals, capital reserve requirements, and interest rate decisions. Ultimately, this strengthens financial stability and overall competitiveness within the financial sector (Baesens et al., 2014).

Traditionally, financial institutions relied on simpler analytical models such as logistic regression to assess credit risk (Hand & Henley, 1997). Consequently, with the increasing volume and complexity of financial data, the tools used to analyse and interpret this information have also evolved (Lessmann et al., 2015). The growing depth of borrower profiles and financial transactions creates the need to evaluate the effectiveness of traditional and modern analytical approaches in today's financial landscape, particularly the introduction of machine learning models that offer enhanced credit risk assessment (Baesens et al., 2014). While these advancements do not replace traditional models, they call for a closer comparison to determine the effectiveness of each technique in managing credit risk (Baesens et al., 2003).

As different credit risk assessment approaches are taken depending on loan type, this study focuses mainly on personal loans in which the risk is dependent on individual financial situations. These loans are frequently used to cover costs such as home improvements, medical expenses, or debt repayment, with amounts typically ranging from a few thousand to tens of thousands in local currency (Abdou & Pointon, 2011). Due to their higher risk level, assessing these loans requires accurate and dependable assessment techniques that consider key factors such as income, credit history, job stability, and debt-to-income ratio to gauge the borrower's ability to repay the loans (Thomas et al., 2017).

This paper examines how traditional techniques like logistic regression compare against more advanced models such as decision trees and random forests in evaluating credit risk for personal loans. The study will discuss the strengths and limitations of each method, factoring in aspects like interpretability, ease of use, predictive accuracy, and adaptability. By comparing these approaches, the research aims to discover the strengths and weaknesses of each technique and the potential synergy in credit risk assessment.

1.2 Research Problem

While traditional credit risk management models, such as logistic regression, have been widely used for decades, they face several challenges in our current data-rich environment. The complexity of current-day financial data, from the volume and scope combined with the need for more accurate risk assessments, exposes the limitations of traditional methods, in particular, handling vast amounts of data and non-linear relationships between borrower profiles and default risk. As a result, traditional methods may be limited in managing the growing complexity of credit risk despite their strength in interpretability.

That said, modern techniques, particularly those based on machine learning, offer significant potential to enhance credit risk assessments by leveraging the full scope of available data. While these modern techniques are increasingly being adopted in banking, there is a gap in understanding the comparative effectiveness against traditional methods. This dissertation aims to address this gap by comparing the analytical models, discussing the strengths and weaknesses of each, and the potential combination of both traditional and modern options in maximising the interpretability and accuracy of credit risk management.

1.3 Research Questions and Objectives

Research Questions:

RQ1. How effective are traditional and modern analytical techniques, and the practical benefits and limitations of each in managing credit risk for personal loans?

RQ2. How can both the traditional and modern techniques be used in unison to maximise the interpretability and accuracy of credit risk management for personal loans?

Research Objectives:

RO1. To evaluate and compare the predictive accuracy and performance of the traditional analytical techniques against the modern analytical techniques.

RO2. To explore the integration of traditional and modern analytical techniques in a hybrid model, assessing the combined use in enhancing both predictive accuracy and interpretability in credit risk management for personal loans.

1.4 Significance of the Study

The research benefits both the academic and financial sectors by comparing traditional and modern methods for credit risk assessment. Academically, it addresses the gaps in research by evaluating how modern methods like decision trees and random forests compare against traditional methods like logistic regression. By assessing the models using different tests and metrics, the research offers a better understanding of the advancements in credit risk assessment, as well as determining which methods are optimal in certain aspects.

For banks and financial institutions, this study aims to promote the combined use of traditional and modern methods in credit risk management. While older methods, such as logistic regression, are easier to understand and interpret, newer ones like decision trees and random forests have higher predictive accuracy. By incorporating the strengths of both, banks can better assess their risk, make better lending decisions, minimise default, and generate financial returns.

1.5 Dissertation Outline

This study is structured into six chapters. Chapter 1 introduces the research background, problem statement, research objectives and questions, as well as the significance of the study, laying a base for the research. Chapter 2 reviews relevant literature on the introduction of regulatory frameworks and analytical techniques, establishing the foundation for model selection and highlighting gaps addressed by this research. Chapter 3 then outlines the research methodology, including the dataset, data preparation, model selection, and evaluation metrics. Based on this, Chapter 4 presents the findings and performance analysis of the logistic regression, decision tree, and random forest models. Chapter 5 further discusses these findings with reference to existing literature and the research objectives, and lastly, Chapter 6 concludes the study by summarising key findings, outlining limitations, and suggesting directions for future research.

2. Literature Review

The literature review is divided into two sections. First, it examines the evolution of credit risk management concerning the Basel Accords. Second, it explores the chronology of analytical techniques, analysing each model in detail to establish a foundation in addressing the research gaps identified in this study.

2.1 Evolution in Response to Regulatory Changes and Digitalisation

The Basel Committee on Banking Supervision (BCBS) introduced Basel I in 1988, standardising credit risk management in the banking sector with the introduction of this framework, requiring banks to maintain a minimum capital adequacy ratio of 8% of their risk-weighted assets (RWAs). To determine RWAs, Basel I categorised assets into broad risk classes, for example, it assigned a 0% risk weight to government securities and a 100% weight to unsecured loans (BIS, 1998). This said, its simplistic risk-weighting methodology failed to account for the increasing complexity of modern banking, leading to weaknesses that became evident, as seen in the Asian Financial Crisis in 1997 (Corsetti, Pesenti and Roubini, 1999).

To overcome these limitations, Basel II was introduced in 2004 with a more refined risk management framework that was structured around three key pillars, which consist of minimum capital requirements, regulatory supervision, and market discipline. A major advancement was the introduction of internal risk models, allowing banks to assess their risk exposure, relying on the accuracy of their own internal data and effective oversight. That said, the 2008 financial crisis exposed significant flaws in Basel II, particularly its inability to prevent banks from underestimating risk. Moreover, the fluctuating capital requirements contributed to financial instability during economic downturns (BIS, 2004).

Following this global financial crisis, Basel III was introduced to tackle the flaws of prior frameworks, minimising the risk of future crises. The main goal of this updated framework was to refine capital requirements through establishing new liquidity regulations, such as the Liquidity Coverage Ratio (LCR) and the Net Stable Funding Ratio (NSFR), which made banks maintain sufficient short and long-term funding. Additionally, it also introduced a leverage ratio to limit excessive borrowing, as well as implemented capital reserves that required banks to set aside funds during periods of economic growth for use during downturns. With that in mind, unlike the prior frameworks, Basel III prioritised systemic stability over individual risk management, which did not directly impact personal loans, which are less affected by systemic financial risks (BIS, 2010).

2.2 Chronology of Analytical Techniques in Credit Risk Management

2.2.1 Traditional Methods

Discriminant analysis was introduced in the mid-20th century, making it one of the earliest statistical techniques used in credit risk management. It works by creating a linear combination of input variables to classify data into distinct groups, such as good and bad, assuming that the relationship between each variable is linear and follows a normal distribution (Altman, 1968).

With this in mind, discriminant analysis has been widely adopted in predicting corporate bankruptcy, with the most notable application being Altman's *Z-score model* (1968). In the research, Altman utilised multiple discriminant analysis (MDA) to combine different financial ratios into a single score, which was then used to assess and categorise companies as either being financially stable or at risk of bankruptcy. This method proved to be highly effective and is still used by banks and investors to assess a company's financial health and predict the likelihood of default.

Despite its effectiveness in predicting corporate bankruptcy, discriminant analysis is not ideally suited for current-day credit risk management as it assumes that the relationships between variables are linear and follow a normal distribution, which may not be an accurate depiction of real-world credit data (Thomas, 2000). Realistically, borrower behaviour is often complex and influenced by many factors that this method cannot fully capture. As the volume and complexity of data have grown, discriminant analysis, alongside prior traditional methods, has become less effective, especially when dealing with large and unstructured datasets.

By the late 20th century, logistic regression emerged as a more flexible alternative to discriminant analysis as it does not require the data to follow a normal distribution. It estimates the probability of a borrower defaulting on a loan based on several variables, such as income, credit history, and loan type. This method works well for binary outcomes like whether a borrower will default or not, making it a good fit for credit scoring (Hosmer & Lemeshow, 2000).

A practical example of logistic regression in the banking sector can be found in Lewis's *An Introduction to Credit Scoring* (1992), which describes how banks use logistic regression to assess loan applicants by analysing factors such as income and debt-to-income ratio. The model assigns a probability score to each applicant, helping lenders classify them as low-risk or high-risk borrowers, enabling financial institutions to make data-driven lending decisions while maintaining consistency in risk assessment.

This said, despite its flexibility, logistic regression still has some drawbacks. It assumes that there is a linear relationship between the variables and the likelihood of default, which may not capture the complex relations between different borrower characteristics (Hand and Henley, 1997). Additionally, logistic regression can be affected by outliers or highly correlated variables, affecting the model's performance and consistency (Lessmann et al., 2015). Even with these limitations, logistic regression is still widely used for its simplicity, ease of interpretation and effectiveness in predicting binary outcomes.

2.2.2 Modern Methods

As financial institutions continued to evolve, the increasing volume and complexity of available data on borrower behaviours highlight the limitations of traditional methods in credit risk management. Traditional approaches relied on assumptions of linearity and simplicity, struggling to adapt to the complexities of modern borrower behaviours. This limitation paved the way for the introduction of more modern, machine learning methods that could handle non-linear relationships and diverse variables.

One of the earlier advancements in machine learning for credit risk assessment was support vector machines (SVMs). Unlike traditional methods like logistic regression, which assume a straightforward, linear relationship between variables, SVMs are comparatively more flexible as they work by finding the best boundary to separate borrowers into different risk categories. The ability to handle complex, non-linear relationships using kernel functions allows for SVMs to pick up on subtle risk factors that simpler models might miss (Cristianini & Shawe-Taylor, 2000).

Research by Lessmann et al. (2015) compared multiple machine learning techniques for credit risk assessment and found that SVMs often outperformed traditional statistical models in terms of predictive accuracy, highlighting SVMs' effectiveness when dealing with high-dimensional data, as they could identify patterns that other models struggled with.

This said, while SVMs provide comparatively enhanced accuracy, they require significant computational power, making them difficult to use on large financial datasets. Moreover, they are not as transparent as the prior methods mentioned, making it harder to explain why the model classifies a borrower as high or low risk, which is a problem for financial institutions requiring justification for their lending decisions to regulators (Danėnas & Garšva, 2009). Due to these challenges, SVMs were less attractive and not widely adopted in mainstream credit risk assessment.

Moving on, another major development was neural networks, working similarly to how the human brain processes information. These models are particularly good at recognising complex patterns in large datasets, making them very effective for predicting credit risk (Goodfellow et al., 2016). Neural networks can process both structured data, such as credit scores and income levels, as well as unstructured data, such as text from loan applications. This allows for the assessment of borrower risk in ways that traditional models are unable to.

Despite their strengths, neural networks have their challenges as they are seen as "black box" models, meaning it is difficult to understand how they arrive at their predictions. This lack of transparency makes them less suitable for regulated industries like banking, where lenders must be able to explain their decisions (Lipton, 2018). Additionally, similar to SVMs, neural networks require large amounts of data and computing power, making them costly to implement compared to simpler models.

As a balance between accuracy and interpretability was sought after, decision trees emerged as a solution as it is able to capture non-linear relationships while minimising computational power. It works by splitting data into smaller groups based on decision rules, making them easy to understand and interpret. In finance, decision trees can classify borrowers based on factors such as credit score and income, effectively capturing complex relationships that traditional methods might overlook (Golbayani, Florescu and Chatterjee, 2020). Their ability to handle multiple influencing factors makes them particularly useful for credit risk assessment.

Additionally, prior research has demonstrated the versatility of decision trees in other fields such as healthcare. For instance, Harper’s work in predicting pregnancy outcomes used decision trees to classify risk factors based on medical data (Harper, 2005). The findings highlighted how decision trees can identify key variables like maternal age and health conditions that influence pregnancy outcomes, providing clear, interpretable results for healthcare providers. While this research is in a different industry, it still proves the effectiveness of decision trees in capturing complex, non-linear relationships within diverse datasets.

However, a weakness of decision trees is that they are prone to overfitting, especially when too many decision rules are applied, causing the model to capture noise instead of meaningful patterns. This leads to inaccurate predictions when applied to new data, making the usage of pruning techniques or ensemble methods like random forests essential to improve generalizability (Zhou and Mentch, 2023).

Building on decision trees, random forests introduced an ensemble approach in the early 21st century. Through the use of combining multiple decision trees and averaging their predictions allows for a significant improvement in accuracy and robustness. Each tree is trained on a random subset of data, enhancing accuracy while reducing overfitting. This approach is especially useful for complex datasets with many variables, as it captures interactions between different factors more effectively than a single decision tree. The method is widely used for credit scoring because it can handle large datasets and provide more reliable predictions (Breiman, 2001).

While random forests are more accurate than decision trees, random forests are harder to interpret. Since the model uses multiple trees, it is difficult to understand exactly how each individual decision tree contributes to the final prediction. Despite this lack of transparency, random forests are very effective for handling complex data and providing more robust credit risk predictions.

3. Methodology

3.1 Dataset Overview

The dataset used in this research is the Statlog German Credit Data, obtained from the UCI Machine Learning Repository. This dataset was chosen for its completeness and relevance in academic research regarding credit risk management. It contains records on 1,000 individuals, with each individual's credit risk assessed based on 20 variables, alongside a target variable indicating credit risk being either good or bad. This dataset contains a wide array of variables that financial institutions would consider, making it suitable for the aim of the research, enabling the comparison between the different analytical techniques.

Table 1: Summary of Dataset Variables

Variable Name	Type	Description
Attribute 1	Categorical	Status of existing checking account
Attribute 2	Numeric	Duration in months
Attribute 3	Categorical	Credit history
Attribute 4	Categorical	Purpose
Attribute 5	Numeric	Credit amount
Attribute 6	Categorical	Savings account/bonds
Attribute 7	Categorical	Present employment since
Attribute 8	Numeric	Instalment rate in percentage of disposable income
Attribute 9	Categorical	Personal status and sex
Attribute 10	Categorical	Other debtors/guarantors
Attribute 11	Numeric	Present residence since
Attribute 12	Categorical	Property
Attribute 13	Numeric	Age
Attribute 14	Categorical	Other instalment plans
Attribute 15	Categorical	Housing
Attribute 16	Numeric	Number of existing credits at this bank
Attribute 17	Categorical	Job
Attribute 18	Numeric	Number of people being liable to provide maintenance for
Attribute 19	Binary	Telephone
Attribute 20	Binary	Foreign worker
Class (Target)	Binary	Good or bad credit risk

The input variables assessed in this dataset include categorical, numerical and binary types. Providing a great opportunity to evaluate each of the analytical techniques' effectiveness in both the processing and interpretation of diverse data types. Table 1 above provides a summary of the dataset's variables.

3.2 Dataset Assumptions

To determine whether the dataset is suitable for evaluation using logistic regression, it is important to check if it aligns with key data requirements. These assumptions include the binary nature of the dependent variable, independence of observations, absence of multicollinearity, linearity in log-odds, and having an adequate sample size. The following sections assess how well the dataset aligns with these assumptions.

First, the dataset defines the target variable credit risk as a binary output, being either good or bad. This is essential for classification-based models, ensuring that the models can effectively differentiate between creditworthy and non-creditworthy applicants.

Another key assumption is the independence of observations, meaning that no repeated measurements or matched data should be present. In this dataset, each row represents a unique loan applicant with distinct financial attributes. Since there are no duplicate records or cases where the same individual is tracked repeatedly, this assumption is satisfied.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
1	(Constant)	1.851	.229		8.081	<.001		
	Attribute1_num	.096	.011	.258	8.540	<.001	.898	1.114
	Attribute2	-.005	.002	-.126	-3.186	.001	.524	1.909
	Attribute3_num	-.028	.014	-.064	-2.046	.041	.826	1.210
	Attribute4_num	.014	.005	.083	2.775	.006	.923	1.084
	Attribute5	-2.143E-5	.000	-.132	-3.095	.002	.450	2.223
	Attribute6_num	.030	.009	.105	3.487	<.001	.908	1.101
	Attribute7_num	.019	.012	.046	1.578	.115	.966	1.035
	Attribute8	-.050	.013	-.123	-3.745	<.001	.763	1.310
	Attribute9_num	.032	.011	.095	3.026	.003	.834	1.198
	Attribute10_num	.000	.030	.000	.004	.996	.948	1.055
	Attribute11	.002	.013	.005	.154	.877	.884	1.131
	Attribute12_num	-.027	.014	-.063	-1.992	.047	.814	1.228
	Attribute13	.002	.001	.059	1.844	.066	.785	1.273
	Attribute14_num	.054	.031	.050	1.726	.085	.982	1.018
	Attribute15_num	-.036	.029	-.042	-1.252	.211	.737	1.357
	Attribute16	-.005	.025	-.007	-.212	.832	.818	1.223
	Attribute17_num	.000	.016	.001	.017	.987	.799	1.252
	Attribute18	-.036	.039	-.028	-.920	.358	.880	1.136
	Attribute19_num	.057	.030	.061	1.896	.058	.792	1.263
	Attribute20_num	-.192	.072	-.079	-2.672	.008	.931	1.075

a. Dependent Variable: ClassTarget_num

Figure 1: VIF Analysis on the Dataset

Moving on to multicollinearity, it affects logistic regression results by inflating coefficient estimates. The VIF analysis, as seen in Figure 1 above, indicates that most variables have low VIF values, suggesting minimal multicollinearity. However, Attribute 5, credit amount (VIF = 2.223) and Attribute 2, duration in months (VIF = 1.909) show moderate correlation, which may slightly affect interpretability. This said, the level of multicollinearity of these variables is still within an acceptable range of regression-based models, with non-linear models like decision trees and random forests remaining unaffected.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Attribute1_BoxTidwell	.300	.036	69.134	1	<.001	1.349	1.257	1.448
	Attribute2_BoxTidwell	-.006	.002	9.655	1	.002	.994	.990	.998
	Attribute3_BoxTidwell	-.063	.038	2.811	1	.094	.939	.872	1.011
	Attribute4_BoxTidwell	.040	.012	11.026	1	<.001	1.041	1.017	1.066
	Attribute5_BoxTidwell	.000	.000	8.463	1	.004	1.000	1.000	1.000
	Attribute6_BoxTidwell	.096	.028	11.960	1	<.001	1.101	1.043	1.163
	Attribute7_BoxTidwell	.063	.035	3.195	1	.074	1.065	.994	1.140
	Attribute8_BoxTidwell	-.145	.040	13.033	1	<.001	.865	.800	.936
	Attribute9_BoxTidwell	.112	.033	11.303	1	<.001	1.119	1.048	1.194
	Attribute10_BoxTidwell	.091	.091	.991	1	.320	1.095	.916	1.310
	Attribute11_BoxTidwell	.018	.038	.213	1	.644	1.018	.944	1.097
	Attribute12_BoxTidwell	-.064	.043	2.186	1	.139	.938	.863	1.021
	Attribute13_BoxTidwell	.005	.002	7.826	1	.005	1.005	1.001	1.008
	Attribute14_BoxTidwell	.213	.116	3.368	1	.066	1.238	.986	1.555
	Attribute15_BoxTidwell	-.035	.086	.163	1	.687	.966	.816	1.144
	Attribute16_BoxTidwell	-.024	.099	.062	1	.804	.976	.804	1.184
	Attribute17_BoxTidwell	.020	.045	.186	1	.666	1.020	.933	1.114
	Attribute18_BoxTidwell	-.137	.167	.672	1	.413	.872	.628	1.210
	Attribute19_BoxTidwell	.226	.129	3.070	1	.080	1.254	.974	1.615
	Attribute20_BoxTidwell	-.508	.287	3.127	1	.077	.602	.343	1.056

a. Variable(s) entered on step 1: Attribute1_BoxTidwell, Attribute2_BoxTidwell, Attribute3_BoxTidwell, Attribute4_BoxTidwell, Attribute5_BoxTidwell, Attribute6_BoxTidwell, Attribute7_BoxTidwell, Attribute8_BoxTidwell, Attribute9_BoxTidwell, Attribute10_BoxTidwell, Attribute11_BoxTidwell, Attribute12_BoxTidwell, Attribute13_BoxTidwell, Attribute14_BoxTidwell, Attribute15_BoxTidwell, Attribute16_BoxTidwell, Attribute17_BoxTidwell, Attribute18_BoxTidwell, Attribute19_BoxTidwell, Attribute20_BoxTidwell.

Figure 2: Box-Tidwell Analysis on the Dataset

Another key assumption is that predictor variables should have a linear relationship with the log-odds of the outcome in logistic regression. The Box-Tidwell test verifies this, as the results in Figure 2 indicate that some variables deviate from this assumption. While this limitation affects logistic regression, decision trees and random forests do not rely on this assumption. This highlights that each technique has its strengths and weaknesses, supporting the idea of combined usage for more comprehensive credit risk assessment.

Finally, a sufficient sample size is best for model stability and reliability. A common rule suggests that at least ten cases are required per predictor for the least frequent outcome. With 1,000 observations and 20 independent variables, the dataset meets the minimum requirement, though a larger sample size is beneficial in improving model stability and generalisation.

3.3 Process Map Overview

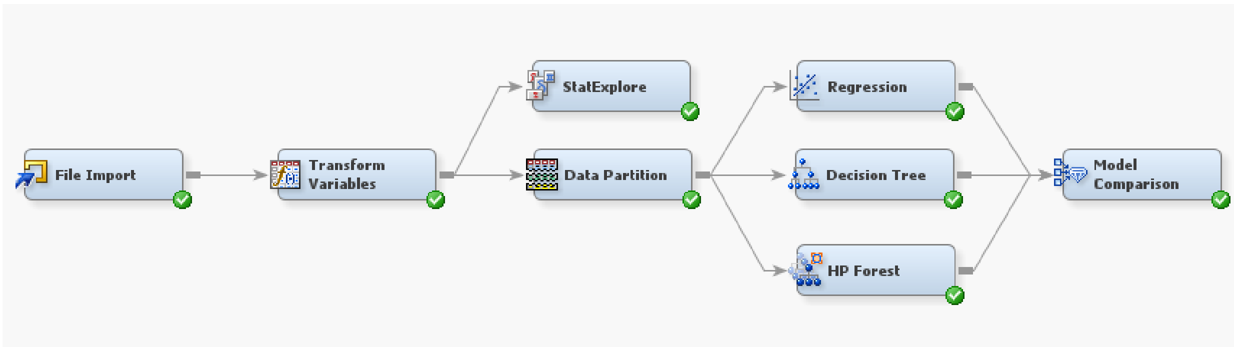

Figure 3: Process Map

This study follows a structured approach to credit risk analysis, ensuring a systematic process for model evaluation. As shown in Figure 3 above, the process begins with importing the raw dataset, followed by data cleaning and variable transformations to correct inconsistencies and to ensure the dataset is standardised for use across different models. After this, the dataset is divided into training and testing subsets to better aid in assessing model performance.

Following data preparation, the predictive models that align with this study are selected and applied to the pre-processed and clean dataset. With this, a final model comparison then evaluates model performance across selected key metrics, determining and presenting the predictive performance of each model relative to one another.

3.4 Data Pre-processing and Cleaning

Data pre-processing and cleaning is a crucial step in ensuring that the dataset is suitable for further analysis, aiming to improve model performance, address inconsistencies and increase predictive accuracy. In the case of this study, missing values found within the dataset, if any, are imputed using mean replacements for numeric variables and mode replacements for categorical variables, ensuring that the overall distribution remains unchanged.

A single version of the dataset is then created in which categorical variables are transformed using one-hot encoding. One-hot encoding allows for the conversion of categorical variables into a series of binary variables, where each is represented by a separate column containing 0s and 1s (Hancock and Khoshgoftaar, 2020). Another function of this encoding method is that it avoids the model assuming an ordinal relationship between categories (James et al., 2013), particularly benefiting the use of logistic regression, ensuring that categorical values are treated separately and not in an ordinal manner. Alternatively, for decision trees and random forests, these techniques natively handle categorical variables, however, using one-hot encoding ensures consistency across all models, allowing for fair and direct comparisons of performance (Loh, 2014).

In addition to the steps mentioned, further data partitioning into training and test sets, 70% and 30% respectively, allows for enhanced model performance evaluation. By splitting the dataset into these subsets, this ensures that the model is trained on a portion of the data and then tested on a separate subset, assessing the model's ability to generalise to new data while minimising overfitting.

3.5 Model Selection

This research focuses solely on evaluating the three analytical techniques, which are logistic regression, decision trees, and random forests. Despite being previously mentioned as having a significant place in the chronology of analytical techniques, discriminant analysis and SVMs are excluded from the selection as they are not ideally suited for this research, with the reasoning for the exclusion being discussed upon earlier.

Focusing on the selected models, logistic regression represents a more traditional analytical technique that is highly interpretable and being able to handle datasets similar to those used in this research, consisting of a mix of categorical, numerical and binary variables, effectively analysing relationships between predictors and the target variable. In the case of this study, logistic regression serves as the benchmark for comparing the performance of modern analytical techniques.

Decision trees, on the other hand, represent the first step in the modernisation of analytical techniques. It works by recursively splitting the dataset into subsets based on the most informative variables, forming a tree structure where each leaf node represents a classification of either being categorised as good or bad credit risk. This said, decision trees are prone to overfitting, affecting the accuracy of the results. This model is included in this study, providing a contrast in terms of interpretability and complexity.

To address the shortcomings of decision trees, random forest, an ensemble method, addresses these issues by combining multiple trees to form a more robust model. Each tree in a random forest is trained on a subset of the data, minimising overfitting while improving accuracy. This technique is well-suited for complex datasets, where interactions between multiple variables are important. The random forest model is included in the comparison to assess whether the enhanced predictive accuracy provides a significant improvement over logistic regression and decision trees, showcasing the potential advantages of ensemble methods in credit risk management.

3.6 Evaluation Metrics

Evaluation metrics play a crucial role in assessing the performance of analytical techniques, helping determine the effectiveness of each model in predicting credit risk. In this study specifically, five key metrics are assessed upon including accuracy, precision, sensitivity, F1-score, and AUC-ROC. Each of these metrics provides valuable insights into each model’s ability to correctly classify credit risks and its overall performance.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Number\ of\ Predictions}$$

Accuracy is defined as the proportion of correct predictions made by the model (Chicco and Jurman, 2020). As shown in the formula above, this metric presents the effectiveness of a model in classifying true positives (good credit risk) and true negatives (bad credit risk). This said, relying on this metric alone may not be sufficient in cases of imbalanced datasets where one class or output occurs more often than the rest, which is why it should be used for assessment in conjunction alongside metrics.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Precision measures the proportion of positive predictions that were indeed correct (Powers, 2020). According to the formula above and in the context of this study, this metric focuses on how accurately each model is in identifying good credit risk, being useful in cases where the cost of false positives or misclassifications is high, as it could potentially lead to financial losses for financial institutions.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Sensitivity measures the proportion of actual positive cases correctly identified by the model (Saito and Rehmsmeier, 2015). This metric is especially important in credit risk assessment as failing to detect a true positive would bring significant revenue losses to financial institutions. Based on the score, having high sensitivity ensures that creditworthy borrowers are correctly classified as low risk, while having low sensitivity means that potentially eligible borrowers are mistakenly classified as high risk, leading to missed loan opportunities, reducing the potential revenue for financial institutions.

$$F1 - Score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

The F1-score serves as a midpoint between precision and sensitivity, providing a balanced measure by combining both metrics (Sasaki, 2007). This metric is particularly useful if an uneven class distribution is present, where focusing on precision or sensitivity alone could lead to inaccurate results. Based on this study, a high F1-score indicates that the model achieves a good balance between correctly identifying low-risk borrowers and minimising the misclassification of high-risk borrowers.

AUC-ROC (Area Under the Receiver Operating Characteristic Curve) measures the model's ability to differentiate between borrowers with good and bad credit risk. It evaluates the balance between the true positive rate (sensitivity) and the false positive rate at different cut-off points (Bradley, 1997). Based on the results produced, a higher AUC indicates better classification performance, with 1 indicating perfect discrimination and 0.5 indicating random guessing.

With this in mind, these metrics are used to evaluate each model’s ability to predict credit risk. By analysing these five metrics together, this study aims to determine the most effective analytical technique for each use case in credit risk management.

4. Findings

4.1 Model Performance Comparison

The performance of the logistic regression, decision tree, and random forest models were evaluated based on key metrics such as accuracy, precision, sensitivity, F1-score and AUC-ROC. The results as shown in Table 2 and Figure 4 below, provide insights into how each model performed in predicting credit risk.

Table 2: Summary of All Performance Metrics

	Logistic Regression	Decision Tree	Random Forest
Accuracy	0.76971214	0.774718398	**0.785982478**
Precision	0.804878048	0.846526655	**0.882165605**
Sensitivity	0.88392851	0.935714285	**0.989285714**
F1-Score	0.84255319	0.888888888	**0.932659932**

4.1.1 Accuracy and Precision Analysis

In terms of accuracy, random forest achieved the highest score at 0.7860, followed by decision trees at 0.7747 and logistic regression at 0.7697. These results indicate that random forest was the most effective model in correctly classifying individuals based on credit risk.

Next, precision indicates the proportion of correctly identified positive cases out of all predicted positives. Consistently, random forest again outperformed the other models with a score of 0.8822, followed by decision trees at 0.8465, and logistic regression with the lowest score at 0.8049. This shows that random forest was the most effective model at correctly identifying high-risk borrowers while minimising false positives.

4.1.2 Sensitivity and F1-Score Analysis

Moving on, sensitivity measures a model's ability to correctly identify actual positive cases. Based on the results, random forest achieved the highest sensitivity at 0.9893, followed by decision trees at 0.9357 and logistic regression at 0.8839. These results highlight that while all models were relatively strong in detecting high-risk borrowers, random forest still outperformed the other models with the lowest number of false negatives, making it the most reliable in capturing true positive cases.

Additionally, F1-score, which is a metric that balances precision and sensitivity, further highlights the performance advantage of the random forest model, with it once again producing the highest score of 0.9327, followed by decision trees at 0.8889 and logistic regression at 0.8426. This further reinforces that random forest was the most effective model in terms of overall classification performance.

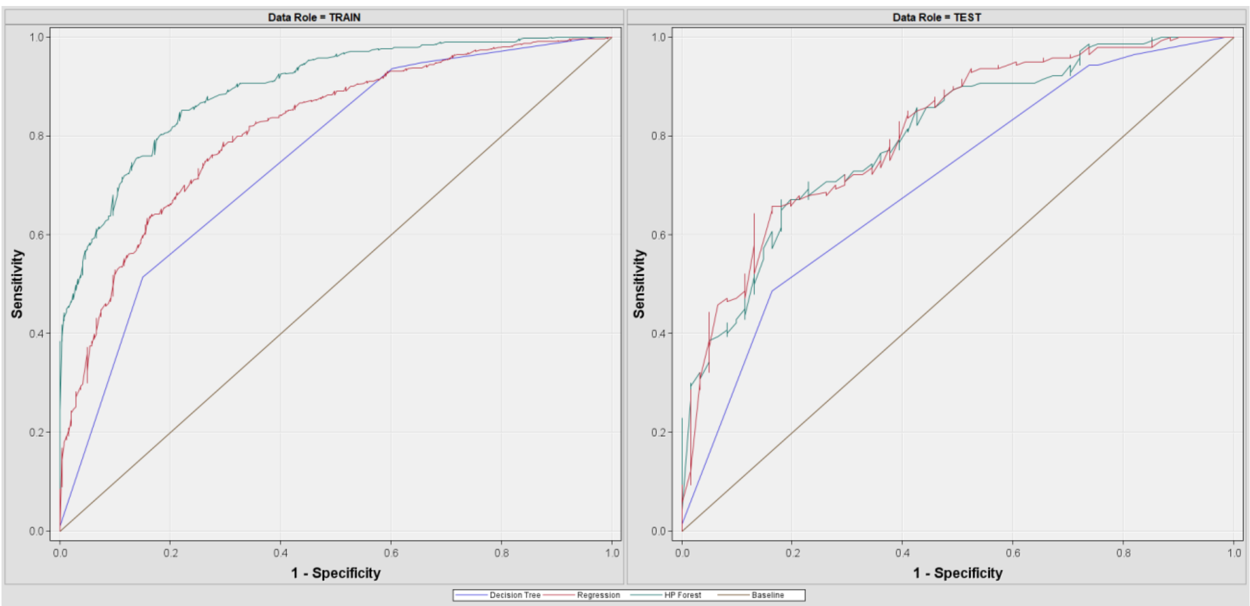

Figure 4: AUC-ROC of All Models

4.1.3 ROC-AUC Analysis

Although random forest outperformed the alternative models in most evaluation metrics, the ROC curve results present a different perspective. In this study, the ROC curve evaluates the trade-off between sensitivity and specificity, highlighting each model's effectiveness in classifying individuals between good and bad credit risk.

The results show that logistic regression performed on par with random forest in terms of classification, while decision trees performed significantly worse. This suggests that while logistic regression underperformed in other metrics, it still demonstrated its relevance in distinguishing between creditworthy and high-risk borrowers.

On the other hand, decision trees performed worse than alternative models in the ROC-AUC metric, despite being considered a modern analytical technique, especially against logistic regression. The weaker performance can be attributed to the model's use of hard classifications, assigning borrowers only as either having good or bad credit risk instead of providing linear, probability-based outputs (Zadrozny and Elkan, 2001). As a result, this heavily reduces both its effectiveness in ROC-AUC analysis and its flexibility in credit risk assessment.

4.2 Individual Model Performance Analysis

Building on the performance comparison of the logistic regression, decision tree, and random forest models, this next subchapter focuses more on each analytical technique's strengths and weaknesses in credit risk assessment.

4.2.1 Logistic Regression

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	Sig.
1	(Constant)	1.851	.229		<.001
	Attribute1_num	.096	.011	.258	<.001
	Attribute2	-.005	.002	-.126	.001
	Attribute3_num	-.028	.014	-.064	.041
	Attribute4_num	.014	.005	.083	.006
	Attribute5	-2.143E-5	.000	-.132	.002
	Attribute6_num	.030	.009	.105	<.001
	Attribute7_num	.019	.012	.046	.115
	Attribute8	-.050	.013	-.123	<.001
	Attribute9_num	.032	.011	.095	.003
	Attribute10_num	.000	.030	.000	.996
	Attribute11	.002	.013	.005	.877
	Attribute12_num	-.027	.014	-.063	.047
	Attribute13	.002	.001	.059	.066
	Attribute14_num	.054	.031	.050	.085
	Attribute15_num	-.036	.029	-.042	.211
	Attribute16	-.005	.025	-.007	.832
	Attribute17_num	.000	.016	.001	.987
	Attribute18	-.036	.039	-.028	.358
	Attribute19_num	.057	.030	.061	.058
	Attribute20_num	-.192	.072	-.079	.008

a. Dependent Variable: ClassTarget_num

Figure 5: Significance Test on Input Variables

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(1.851+0.096X_1-0.005X_2-0.028X_3+0.014X_4-2.143\times10^{-5}X_5+0.030X_6-0.050X_8+0.032X_9-0.192X_{20})}}$$

Y = Risk Classification

X_1 = Status of existing checking account

X_2 = Duration in months

X_3 = Credit history

X_4 = Purpose

X_5 = Credit amount

X_6 = Savings account/bonds

X_8 = Instalment rate in percentage of disposable income

X_9 = Personal status and sex

X_{20} = Foreign worker

The logistic regression equation above was derived by running a test of significance as seen in Figure 5 above, in which input variables with p-values below 0.05 were considered statistically significant (Hosmer and Lemeshow, 2000), retaining only the 9 significant variables which affect the output for credit risk prediction, reducing the risk of overfitting while maintaining interpretability. By focusing on statistically significant borrower attributes, the model effectively captures key financial indicators that determine a borrower's likelihood of default. With the equation established, the model determines a borrower’s credit risk by assigning a probability score between 0 and 1, representing their estimated risk level. Borrowers with scores below 0.5 are classified as having bad credit risk (0), while those scored above 0.5 are categorised as having good credit risk (1). Unlike rigid classification rules in alternative models, this probability-based approach provides better flexibility in assessing credit risk (Hand and Henley, 1997).

This said, compared to alternative models, the main advantage of logistic regression is in its ability to produce a clear mathematical equation for interpretation (Hosmer and Lemeshow, 2000). Based on the equation above, each coefficient reflects a variable’s influence on credit risk, with positive values indicating higher default likelihood and negative values suggesting lower risk. For example, holding other variables constant, variable X_1 (Status of existing checking account) has the highest positive coefficient at +0.096, indicating that having a stronger checking account status is the most important factor in reducing credit risk compared to other significant variables selected. In contrast, holding other variables constant, variable X_{20} (Foreign worker) has the most negative coefficient at -0.192, which shows that being a foreign worker is strongly related to an increased credit risk.

The enhanced interpretability is particularly valuable in financial regulation and decision-making as it allows financial institutions to understand how specific borrower attributes impact credit risk (Thomas, 2000). Compared to more complex models, logistic regression provides straightforward probability estimates, enhancing the interpretation and justification of lending decisions such as loan approvals or rejections (Lipton, 2018).

Despite the strength in interpretability, logistic regression performs comparatively worse against more advanced and modern models in predictive performance. As discussed in Chapter 4.1, it performed the weakest across key evaluation metrics, including accuracy, precision, sensitivity, and F1-score (Baesens et al., 2003). One major limitation is the model's assumption of a linear relationship between predictors and the outcome, which restricts its ability to capture complex, non-linear patterns in financial data (James et al., 2013). Additionally, it assumes that predictor variables are independent, which does not reflect real-world credit risk data (Abdou and Pointon, 2011). These weaknesses aside, logistic regression remains valuable due to its transparent probability-based classification, which provides enhanced interpretability, making it particularly suitable for financial decision making where clear and justifiable credit risk assessments are an essential part of the process (Thomas, Crook and Edelman, 2017).

4.2.2 Decision Tree

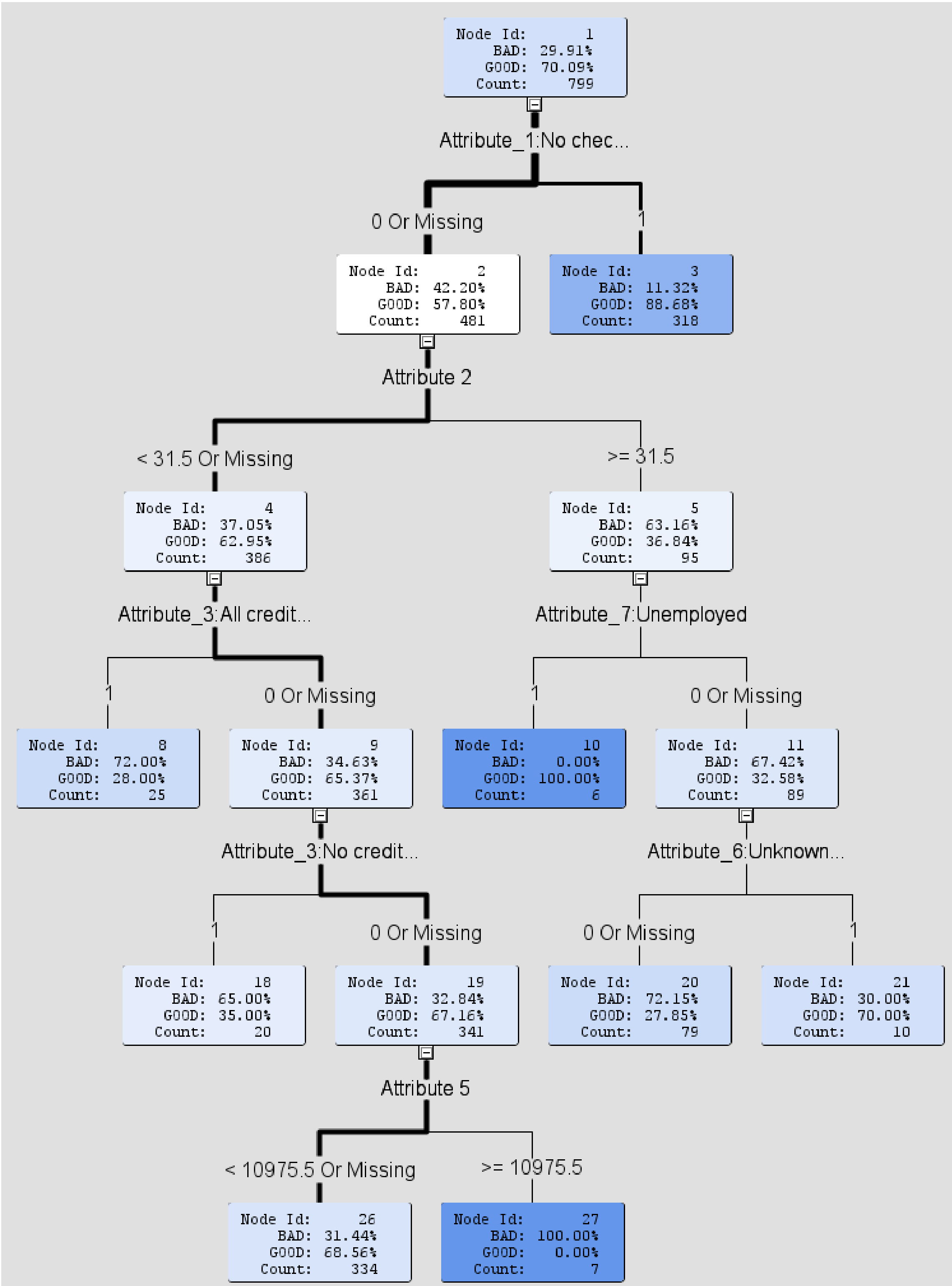

Figure 6: Decision Tree

Attribute 1 = Status of existing checking account

Attribute 2 = Duration in months

Attribute 3 = Credit history

Attribute 5 = Credit amount

Attribute 6 = Savings account/bonds

Attribute 7 = Present employment since

Moving on to decision trees, this model provides a more structured approach to evaluate credit risk based on borrower attributes. As shown in Figure 6, the tree begins with Attribute 1 as it is the most significant variable in predicting default risk, with borrowers without a checking account being categorised as high-risk due to the lack of financial reserves (Hofmann, 1994). After this initial split, the next variable that the model assesses is Attribute 2, as loan duration often directly affects the probability of default, with shorter repayment periods normally indicating a lower likelihood of default (Thomas, Crook and Edelman, 2017).

Going further down the tree, Attribute 3 is then considered, as individuals with a strong credit history are more likely to be categorised as low risk, whereas individuals with no credit history require further assessment (Abdou and Pointon, 2011). The model then revisits Attribute 3 again at a deeper level, emphasising the attribute's significance in assessing creditworthiness. If credit history is unavailable, the decision tree shifts its focus instead to Attribute 5, assessing the individual's financial responsibility. Additional factors, such as Attribute 7 and Attribute 6, further refine the classification by incorporating income stability and financial reserves into the risk assessment process (Thomas, 2000).

Based on the example above, the primary advantages of decision trees, as presented above, are the model's transparency and ease of interpretation. This model provides a concise and hierarchical structure that makes it easy to understand how credit risk decisions are made. By ranking variables based on their contribution to reducing uncertainty, the model highlights the most influential features in a tree-like visualisation, signifying and arranging each chosen variable in a descending order of importance with respect to the target outcome (Lessmann et al., 2015). This is a feature that is highly sought after in financial and regulatory settings, as justification behind credit risk assessments is essential. Furthermore, decision trees can also process both numerical and categorical data, making it highly adaptable to various types of data (James et al., 2013). Their ability to automatically prioritise the most relevant variables also simplifies the variable selection process, ensuring that only the most significant predictors influence the outcome. Additionally, the model requires minimal data preprocessing since they do not rely on assumptions about linear relationships between variables, making it particularly useful for financial datasets, where borrower attributes often correlate in complex and non-linear ways (Baesens et al., 2003).

Despite having these benefits, decision trees have several limitations, with the main weakness being overfitting, which occurs when the model becomes too tailored to the training data, reducing its effectiveness on test data (Loh, 2014). Additionally, with the classification output being based only on a single decision tree, this further restricts its ability to capture complex financial relationships, potentially oversimplifying interactions between variables (Zhou and Mentch, 2023). Referring back to Chapter 4.1, although the decision tree model performed well in terms of sensitivity (0.9357), the comparatively worse accuracy (0.7747) and precision (0.8465) results indicate subpar predictive performance compared to alternative models. Moving on, this model too had the worst AUC-ROC performance due to the model's tendency to make definite output categorisations instead of providing probability scores, such as categorising the credit risk of a borrower as a definite good or bad, lacking the flexibility to assess credit risk based on a scale (Baesens et al., 2014). These limitations aside, its rule-based classification remains particularly valuable in credit risk assessment, where clarity and interpretability in decision making are essential (Thomas, Crook and Edelman, 2017).

4.2.3 Random Forest

Variable Name	Number of Splitting Rules	Train: Gini Reduction	Train: Margin Reduction	OOB: Gini Reduction	OOB: Margin Reduction	Label
Attribute 1	53	0.025172	0.050344	0.02440	0.04872	Attribute 1
Attribute 3	37	0.008106	0.016211	0.00450	0.01340	Attribute 3
Attribute 2	27	0.006101	0.012203	0.00152	0.00752	Attribute 2
Attribute 13	18	0.002397	0.004793	-0.00073	0.00169	Attribute 13
Attribute 4	14	0.002346	0.004692	-0.00025	0.00286	Attribute 4
Attribute 5	12	0.002636	0.005271	-0.00067	0.00209	Attribute 5
Attribute 15	11	0.001391	0.002781	0.00021	0.00168	Attribute 15
Attribute 12	9	0.001377	0.002755	0.00064	0.00197	Attribute 12
Attribute 6	9	0.001645	0.003290	0.00056	0.00142	Attribute 6
Attribute 16	8	0.000567	0.001135	-0.00067	-0.00025	Attribute 16
Attribute 8	8	0.000906	0.001812	-0.00003	0.00054	Attribute 8
Attribute 14	7	0.000494	0.000988	0.00013	0.00094	Attribute 14
Attribute 19	6	0.000369	0.000738	-0.00039	-0.00006	Attribute 19
Attribute 11	4	0.000313	0.000627	-0.00024	0.00007	Attribute 11
Attribute 17	3	0.000200	0.000399	-0.00005	0.00007	Attribute 17
Attribute 7	3	0.000344	0.000689	-0.00006	0.00062	Attribute 7
Attribute 9	3	0.000273	0.000547	-0.00028	0.00021	Attribute 9
Attribute 10	2	0.000147	0.000294	0.00022	0.00038	Attribute 10
Attribute 18	2	0.000110	0.000220	-0.00022	-0.00010	Attribute 18
Attribute 20	2	0.000130	0.000259	0.00006	0.00009	Attribute 20

Figure 7: Variable Importance Results (RF)

Attribute 1 = Status of existing checking account

Attribute 2 = Duration in months

Attribute 3 = Credit history

Attribute 4 = Purpose

Attribute 5 = Credit amount

Attribute 6 = Savings account/bonds

Attribute 7 = Present employment since

Attribute 8 = Instalment rate in percentage of disposable income

Attribute 9 = Personal status and sex

Attribute 10 = Other debtors/guarantors

Attribute 11 = Present residence since

Attribute 12 = Property

Attribute 13 = Age

Attribute 14 = Other instalment plans

Attribute 15 = Housing

Attribute 16 = Number of existing credits at this bank

Attribute 17 = Job

Attribute 18 = Number of people being liable to provide maintenance for

Attribute 19 = Telephone

Attribute 20 = Foreign worker

Random forest improves upon the flaws of decision trees by combining multiple trees to improve classification accuracy and reduce overfitting. Unlike decision trees, which utilise only a single tree, random forest builds multiple trees using different data subsets and features, increasing predictive performance (Breiman, 2001). Referring back to Table 2, the random forest model outperformed the logistic regression and decision tree models in accuracy, precision, sensitivity, and F1-score. This enhanced performance comes from its ability to capture complex, non-linear relationships, as well as minimise individual tree biases, leading to more robust and reliable predictions (Lessmann et al., 2015).

Despite its superior predictive performance, the model is known as a “black box” model, making it less transparent than alternative models in terms of interpretability, as logistic regression provides clear coefficients (Hosmer and Lemeshow, 2000), while decision trees provide transparent, rule-based classifications (Loh, 2014). As shown in Figure 7, this illustrates the variable importance rankings based on the number of splitting rules in the random forest model, a different approach compared to the t-test and coefficient-based logistic regression model and the hierarchical, rule-based approach of decision trees. While the random forest model’s approach highlights the most influential features in the prediction process, it does not provide any clear explanation as to how or why these variables affected the final output. This reinforces the “black box” nature of the model, where the model’s strong performance comes at the cost of a lack of interpretation (Lipton, 2018). This limitation presents a challenge in regulatory and financial settings where decision transparency is essential. Moreover, its high computational demands and longer processing times make it less practical for real-time credit scoring applications (Baesens et al., 2014).

In summary, the random forest model displayed the best predictive performance but at the expense of interpretability and computational efficiency (James et al., 2013). While the ability to capture complex, non-linear relationships and minimise individual tree biases makes the model highly effective for credit risk assessment, the lack of transparency may limit its applicability in processes that require clear explanations and backing for financial decision making (Baesens et al., 2014).

5. Discussion

5.1 Effectiveness of Traditional vs. Modern Techniques in Credit Risk Management (RQ1)

Based on the findings in Chapter 4 above, each analytical technique offers unique advantages and limitations in credit risk assessment. Starting with logistic regression, this model is highly regarded for its interpretability, as it provides a clear mathematical equation with coefficients that represent the influence of each borrower attribute on determining credit risk (Hosmer & Lemeshow, 2000). Additionally, the transparency of logistic regression enables financial institutions to justify lending decisions and present risk assessments to stakeholders in an easily interpretable manner. Despite this, the model's predictive ability is comparatively weaker compared to the more modern machine learning models, as it assumes that the predictor variables are independent and that the relationship with credit risk is linear (Hand & Henley, 1997). These assumptions limit the model's ability to capture more complex patterns, especially as the financial data available continues to expand in volume and complexity (Abdou & Pointon, 2011).

On the other hand, machine learning approaches like the decision tree and random forest models offer a significant improvement in prediction accuracy. For instance, decision trees, utilising a structured and rule-based classification system, enhance the interpretation of credit risk assessment decisions by outlining the hierarchical importance of the selected input variables (Breiman, 2001). This said, despite its transparency and interpretability, decision trees are prone to overfitting, meaning that the model becomes too tailored to the training data, resulting in subpar performance on the test data (Loh, 2014). Additionally, decision trees output definite classifications, which in this study, a more flexible, scale-based classification would be more effective in assessing credit risk, resulting in the comparatively weaker performance in the AUC-ROC metric (Saito & Rehmsmeier, 2015).

Based on the results of this study, the random forest model resulted in the highest predictive performance, outperforming both the logistic regression and decision tree models across all key evaluation metrics, including accuracy, precision, sensitivity, and F1-score. This improved performance can be attributed to the model's ensemble nature, combining multiple decision trees to enhance reliability and reduce bias (Breiman, 2001). However, a major disadvantage of this model is its lack of transparency. Unlike logistic regression, which clearly defines how each variable directly influences the output, or decision trees, which produce straightforward classification rules, random forest functions as a "black box" model, providing highly accurate predictions but at the expense of transparency and interpretability on prediction justification. This lack of interpretability can be problematic in financial environments, where justification behind the decision-making processes is required (Lipton, 2018). Moreover, this model requires greater computational resources and longer processing times, which makes it less practical for real-time credit scoring applications (Golbayani et al., 2020).

In summary, each model comes with its own set of strengths and weaknesses. Traditional analytical techniques like logistic regression, despite having comparatively lower performance, remain valuable due to their clarity and ease of interpretation, while modern machine learning techniques offer significant improvements in prediction performance, with the caveat being the subpar interpretability and computational complexity (Thomas, Crook & Edelman, 2017). This said, the strengths and limitations of each analytical model must be carefully considered by financial institutions when selecting the best approach for credit risk assessment.

5.2 Combining Traditional and Modern Techniques for Improved Credit Risk Assessment (RQ2)

As both traditional and machine learning models have strengths in different aspects, the combined use of both allows for an enhanced approach to credit risk assessment. Through the integration of this, financial institutions have the opportunity to develop models that maximise both predictive accuracy and interpretability.

One approach to achieve this is by using logistic regression first for feature selection. By identifying and retaining only the significant borrower attributes, this approach reduces the complexity and ensures only key variables are retained and considered (James et al., 2013). Once relevant variables are selected, a more advanced machine learning technique like random forest can then be applied to classify borrowers' credit risk. This combined approach could allow for more transparent variable selection justification while maximising predictive accuracy.

Next, another effective approach is to incorporate decision trees to act as an intermediary between the logistic regression and random forest models. As decision trees offer a hierarchical, rule-based system, the model provides an additional layer of interpretability before applying a more complex ensemble method like random forest. This hybrid approach allows financial institutions to gain insights into key factors influencing credit risk while still benefiting from the superior predictive performance of machine learning methods.

Additionally, the ROC-AUC results in Figure 4 suggest that, despite having comparatively lower predictive performance, logistic regression can still serve as a useful benchmark model, with its classification performance being on par with random forest, indicating its ability to produce consistent and reliable predictions (Bradley, 1997). This allows for financial institutions to incorporate logistic regression as a benchmark, ensuring that the predictions produced from the modern machine learning models align with the already existing credit risk assessment frameworks in the financial sector.

Ultimately, the integration of both traditional and modern analytical methods offers numerous benefits, leading to a more effective credit risk assessment approach. By leveraging the traditional models' strengths in transparency and regulatory compliance alongside the modern models' enhanced predictive performance, financial institutions can develop models that are more accurate while providing clear and justifiable interpretations, effectively improving operations while minimising the limitations.

6. Conclusion

This study evaluated the predictive performance of the logistic regression, decision trees, and random forests models in credit risk assessment. While logistic regression offers strong interpretability, the model's predictive performance is subpar compared to more advanced models. Decision trees, on the other hand, despite offering structured, rule-based classification, suffer from overfitting and weaker classification performance. Lastly, despite being a "black-box" model with limited interpretability, the random forest model produced the highest predictive performance, outperforming both the logistic regression and decision tree models in key metrics such as accuracy, precision, sensitivity, and F1-score.

Based on the findings of this study, by leveraging both the traditional and modern techniques, the combined use of these analytical models could significantly enhance credit risk assessment. Through careful selection and integration of each technique, the integration of a hybrid approach allows financial institutions to maximise predictive performance and interpretability, which is highly beneficial in meeting regulatory requirements and making well-informed decisions.

This said, this study has several limitations. Firstly, the analysis was done on a relatively small and limited dataset, which may not fully capture the complexity of real-world credit risk data. A larger, more representative dataset could improve the reliability and accuracy of the findings. Additionally, only three analytical models were selected and examined, while other advanced techniques such as gradient boosting, XGBoost, or neural networks were not considered. Finally, the lack of extensive model hyperparameter tuning could have limited the optimisation of model performance, potentially hindering the selected analytical methods from achieving their best possible performance.

In summary, future research should address these limitations by applying a broader range of models, using larger and more diverse datasets, and exploring additional model tuning and validation techniques. Refining these hybrid approaches could lead to more accurate and interpretable predictions, further improving credit risk management processes in the near future.

References

1. Abdou, H.A. and Pointon, J., 2011. Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management*, 18(2–3), pp.59–88.
2. Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), pp.589–609.
3. Baesens, B., Van Gestel, T., Vanthienen, J., Stepanova, M., & Suykens, J. A. K., 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), pp. 627-635.
4. Baesens, B., Van Vlasselaer, V. and Verbeke, W., 2014. *Analytics in a Big Data World: The Essential Guide to Data Science and its Applications*. Hoboken: Wiley.
5. Bank for International Settlements (BIS), 1998. *International convergence of capital measurement and capital standards: A revised framework*. Basel: BIS.
6. Bank for International Settlements (BIS), 2004. *International convergence of capital measurement and capital standards: A revised framework*. Basel: BIS.
7. Bank for International Settlements (BIS), 2010. *Basel III: A global regulatory framework for more resilient banks and banking systems*. Basel: BIS.
8. Bank for International Settlements (BIS), 2011. *Basel III: A global regulatory framework for more resilient banks and banking systems*. Basel: BIS.
9. Box, G.E.P. and Tidwell, P.W., 1962. Transformation of the independent variables. *Technometrics*, 4(4), pp.531–550.
10. Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, *30*(7), pp.1145-1159.
11. Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5–32.
12. Chicco, D. and Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*, pp.1-13.
13. Corsetti, G., Pesenti, P. and Roubini, N., 1999. What caused the Asian currency and financial crisis? *Japan and the world economy*, *11*(3), pp.305-373.
14. Cristianini, N. and Shawe-Taylor, J., 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press.

15. Danėnas, P. and Garšva, J., 2009. Support vector machines and their applications in credit risk. *International Journal of Management and Applied Science*, 4(2), pp.1-4.
16. Golbayani, P., Florescu, I. and Chatterjee, R., 2020. A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. *The North American Journal of Economics and Finance*, 54, p.101251. Available at: https://doi.org/10.1016/j.najef.2020.101251 [Accessed 6 Apr. 2025].
17. Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep Learning*. Cambridge, MA: MIT Press.
18. Hand, D.J. and 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), pp.523–541.
19. Hancock, J.T. and Khoshgoftaar, T.M., 2020. Survey on categorical feature encoding techniques for machine learning classifiers. *Journal of Big Data*, 7(1), pp.1–22.
20. Harper, P.R., 2005. A review and comparison of classification algorithms for medical decision making. *Health Care Management Science*, 8(2), pp.91–105.
21. Hofmann, H., 1994. *Statlog (German Credit Data)* [Dataset]. UCI Machine Learning Repository. Available at: https://doi.org/10.24432/C5NC77 [Accessed 16 Feb. 2025].
22. Hosmer, D.W. and Lemeshow, S., 2000. *Applied Logistic Regression*. 2nd ed. New York: Wiley.
23. James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer.
24. Lessmann, S., Baesens, B., Seow, H.V. and Thomas, L.C., 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), pp.124–136.
25. Lewis, C., 1992. *An Introduction to Credit Scoring*. San Rafael: Athena Press.
26. Lipton, Z.C., 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), pp.31-57.
27. Loh, W.Y., 2014. Fifty years of classification and regression trees. *International Statistical Review*, 82(3), pp.329–348.
28. O'Brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), pp.673–690.

29. Powers, D.M.W., 2020. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *International Journal of Machine Learning and Computing*, 10(1), pp.1–19. doi:10.18178/ijmlc.2020.10.1.511.
30. Saito, T. and Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, *10*(3), p.e0118432.
31. Sasaki, Y., 2007. The truth of the F-measure. *Teach tutor mater*, *1*(5), pp.1-5.
32. Thomas, L.C., 2000. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International journal of forecasting*, *16*(2), pp.149-172.
33. Thomas, L., Crook, J. and Edelman, D., 2017. *Credit scoring and its applications*. Society for Industrial and Applied Mathematics.
34. Zadrozny, B. and Elkan, C., 2001, June. Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. In *Icml* (Vol. 1, No. 05).
35. Zhou, S. and Mentch, L., 2023. Trees, forests, chickens, and eggs: when and why to prune trees in a random forest. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *16*(1), pp.45-64.

Appendix

Appendix 1: Table of Fit Statistics (LR)

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Class Target	Class (Target)	AIC	Akaike's Information Criterion	784.9621	.	.
Class Target	Class (Target)	ASE	Average Squared Error	0.154123	.	0.155924
Class Target	Class (Target)	AVERR	Average Error Function	0.469939	.	0.473767
Class Target	Class (Target)	DFE	Degrees of Freedom for Error	782	.	.
Class Target	Class (Target)	DFM	Model Degrees of Freedom	17	.	.
Class Target	Class (Target)	DFT	Total Degrees of Freedom	799	.	.
Class Target	Class (Target)	DIV	Divisor for ASE	1598	.	402
Class Target	Class (Target)	ERR	Error Function	750.9621	.	190.4544
Class Target	Class (Target)	FPE	Final Prediction Error	0.160823	.	.
Class Target	Class (Target)	MAX	Maximum Absolute Error	0.966585	.	0.964255
Class Target	Class (Target)	MSE	Mean Square Error	0.157473	.	0.155924
Class Target	Class (Target)	NOBS	Sum of Frequencies	799	.	201
Class Target	Class (Target)	NW	Number of Estimate Weights	17	.	.
Class Target	Class (Target)	RASE	Root Average Sum of Squares	0.392584	.	0.394872
Class Target	Class (Target)	RFPE	Root Final Prediction Error	0.401028	.	.
Class Target	Class (Target)	RMSE	Root Mean Squared Error	0.396829	.	0.394872
Class Target	Class (Target)	SBC	Schwarz's Bayesian Criterion	864.5792	.	.
Class Target	Class (Target)	SSE	Sum of Squared Errors	246.2878	.	62.6815
Class Target	Class (Target)	SUMW	Sum of Case Weights Times Freq	1598	.	402
Class Target	Class (Target)	MISC	Misclassification Rate	0.230288	.	0.218905

Appendix 2: Dissertation Submission Form

NOTTINGHAM UNIVERSITY BUSINESS SCHOOL
UNDERGRADUATE, MASTERS, AND MBA PROGRAMMES
ETHICS REVIEW FORM
(To be submitted with the Dissertation/Management Project Submission Form)

Please complete the below and return this form when you submit your dissertation/management project.

SECTION 1 ***Student to Complete***

Last Name **Sia**	**First Name** **Tung Yee**
Student ID **20511127**	**Degree** **BSc (Hons) Finance, Management and Business Analytics**
Dissertation/Management Project Title **The Impact of Analytical Techniques in Credit Risk Management**	
Supervisors Professor Seow Hsin-Vonn	**Submission Date** **17 April 2025**

Signed Elis ______________________________

I confirm that the attached dissertation/management project is my own work. If ethical approval was required for the project, I confirm that I carried out the research in accordance with a pre-approved protocol or in accordance with the protocol agreed by the NUBS Research Ethics Committee.

SECTION 2 ***Student to Complete***

a) This information is required to help us complete information about your research: Have you spent any time overseas (i.e., apart from in Malaysia) undertaking research relating to your dissertation/management project, e.g., collecting data? *(Please tick)*

YES ☐ NO ☑

If 'yes' in which countries did you complete your research in:

………

How long did you spend in the above countries in order to complete your research? ***(Even if it is your home country please specify the research time period):***

………

b) Is your research based within a company (i.e., either a placement Project/Internship organised by the Business School or alternatively arranged by yourself)

YES ☐ NO ☑

If 'yes' please state the name of the company and the country in which it was based.

………

c) **Did your research require ethics review?** Yes / No *(strike through irrelevant option)* Note: If you used secondary data, your research did not require ethics review.

If yes, please state the name of the pre-approved research protocol that you used, or enter the reference number for the individual project approval.

- ☐ Pre-approved protocol: business plan, company project, interviews, internet survey *(strike through irrelevant options)*

- ☐ Individual project review: Reference number _________________

SECTION 3 ***Office use only***

Date Received	Received after Deadline Yes / No Does Student have Extenuating Circumstance Yes / No Is evidence attached Yes / No

Please return this form at the same time you submit your dissertation.