Forecasting Customer Lifetime Value through Risk Prediction: An Explainable Machine Learning Approach for the Telecommunication Industry

Master Business Information Technology Thesis
August 2023

Author:

Edo Belva Firmansyah | Faculty of EEMCS, University of Twente

Supervisors:

Dr. M. R. Machado | Faculty of BMS, University of Twente

Dr. João Luiz Rebelo Moreira | Faculty of EEMCS, University of Twente

UNIVERSITY OF TWENTE.

Abstract

As businesses increasingly prioritize understanding customers' value through metrics such as Customer Lifetime Value (CLV), they leverage it to guide customer engagement and retention strategies. However, a systematic literature review revealed a noticeable gap: the limited integration of customers' risk factors into CLV calculations. This gap becomes even more pronounced in the telecommunications sector. Despite being rich in detailed customer data, the sector remains largely uncharted territory for risk-adjusted CLV predictions via machine learning (ML). This thesis aims to bridge this disparity, emphasizing the prediction of risk-adjusted CLV in the non-contractual (B2C) setting of the telecommunications industry using ML techniques.

This thesis presents a novel approach to integrate customers' risk into CLV calculations in the telecommunications industry by introducing a Risk-Adjusted Return (RAR) metric. The thesis employed the Design Science Research Methodology (DSRM) to develop this research and utilized the Cross-Industry Standard Process for Data Mining (CRISP-DM) to construct ML models. The customers' risk involved in RAR calculation incorporated the probability of customer churn and beta value, modifying the discount rate used in CLV calculation to reflect risk. Four different approaches were proposed to calculate RAR. ML models, including Logistic Regression, XGBoost, CatBoost, and Random Forest were built to predict both customer churn probability and RAR. To validate these models, eXplainable AI (XAI) techniques, such as feature importance, SHAP global explanation, SHAP local explanation, and LIME, were utilized.

The results indicated statistical differences among the four proposed RAR calculation approaches, validating their distinctiveness. The churn model demonstrated an accuracy of 85%, while the RAR models exhibited robust performance with an R^2 of 0.92 and Mean Absolute Percentage Error (MAPE) around 20%. Among the algorithms, XGBoost emerged as the best model for churn prediction, while CatBoost outperformed others in predicting RAR for all four approaches. The most influential features for RAR across the approaches were found to be the number of loyalty point acquired by the customer, average revenue from customer in the last 3 months including with it standard deviation, the total revenue from the customer along with the standard deviation, the probability of customers' churn and the beta value. These features align with the features used in traditional RAR calculations thus increase the model validity.

The thesis concludes by asserting the statistical significance of the proposed RAR and the robustness of the proposed models, with their feature importance aligning with traditional calculation models. For future research, there are opportunities to explore different risks, decompose the revenue components for individual RAR calculations, utilize more advanced ML algorithms and hyperparameter tuning, and further incorporate ML in XAI.

Keywords: Risk Adjusted Revenue, Machine Learning, explainable AI

Table of Contents

List of Figures

List of Tables

List of Abbreviation

AB	Adaptive Boosting
AI	Artificial Intelegence
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AUC-ROC	Area Under the Receiver Operating Characteristic curve
B2B	Business-to-Business
B2C	Business-to-Consumer
CAPM	Capital Asset Pricing Model
CLV	Customer Lifetime Value
CRISP-DM	Cross-Industry Standard Process for Data Mining
CRM	Customer Relationship Management
CV	Cross-Validation
DARPA	Defence Advanced Research Projects Agency
DEA	Data Envelopment Analysis
DNN	Deep Neural Network
DSRM	Design Science Research Methodology
DT	Decision Tree
FSI	Financial Services Industry
GB	Gradient Boosting
GDPR	General Data Protection Regulation
HMM	Hidden Markov Model
LIME	Local Interpretable Model-agnostic Explanations
LOS	Length of Stay
LR	Logistic regression
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MDP	Markov Decision Process
ML	Machine Learning
MPT	Modern Portfolio Theory
MSE	Mean Squared Error
MSR	Multiple Source of Risk
NPV	Net Present Value
OHE	One-Hot Encoding
PD	Probability of Default
PST	Portfolio Selection Theory
RALTV	Risk-Adjusted Lifetime Value
RARFM	Risk-Adjusted Recency Frequency and Monetary
RF	Random Forest
RFM	Recency, Frequency, Monetary
RM	Relationship Marketing
RMSE	Root Mean Squared Error
RQ	Research Question

SHAP	SHapley Additive exPlanations
SLR	Systematic Literature Review
SQ	Sub-Questions
SVM	Support Vector Machine
WACC	Weighted Average Cost of Capital
WOE	Weight of Evidence
WoS	Web of Science
XAI	Explainable AI
XGBoost	Extreme Gradient Boosting

1 Introduction

1.1 Background Information

Customers are the foundation of any business. They serve as the main source of revenue and simultaneously act as a valuable asset that influences the company's performance, profitability, and growth. This understanding has prompted a shift in corporate practices towards a more customer-centric approach, leading to the emergence of Relationship Marketing (RM) in the 1980s. Relationship Marketing aims to attract, retain, and enhance customer relationships (Ryals & Knox, 2005). It is important for the company to know the value of its customers and develop appropriate strategies to retain them. The most widely used metric to measure the customers' value to the company is Customer Lifetime Value (CLV) (Kumar & Reinartz, 2016; Glady et al., 2015; Gupta et al., 2004). CLV represents the total net present value of all future cash flows expected to be generated by a customer over the course of their relationship with a company, minus the costs associated with acquiring and serving that customer (Kumar & Reinartz, 2016; Gupta et al., 2004).

By considering the future value of a customer, a company can make better decisions about how to treat its customers and decide which customers are worth retaining and which ones to let go. Furthermore, by measuring the CLV, companies are also able to improve company profitability by selecting more profitable customers, targeting them effectively, and focusing on long-term customer relationships (Dahana et al., 2019; Kumar & Reinartz, 2016; Qi et al., 2012). Identifying the most valuable customers allows a company to focus its resources on retaining them and increasing their value. This can be accomplished through targeted marketing campaigns (Benedek et al., 2014), loyalty programs (Kang et al., 2015), and personalized offers (Daqar & Smoudy, 2019). By improving customer retention and increasing the value of its customer base, a company can improve its profitability and long-term success. Therefore, CLV has become a critical aspect of RM, serving as a significant metric for developing tailored marketing and retention strategies for each customer (Dahana et al., 2019; Kumar & Reinartz, 2016; Qi et al., 2012).

Moreover, CLV is not just a measure of customers' value but can also be used to measure a company's financial performance and estimate its valuation (Gupta et al., 2004; Hogan et al., 2002). Combining CLV with projected total customer growth helps estimate the company's current value. To the extent that the customer base forms a significant part of a company's overall value, CLV can serve as a useful proxy for firm value (Gupta et al., 2004). The total customers' value of a company is aligned with the total Market Capital of the firm (Rust et al., 2004), hence a company with a higher CLV tends to have a higher stock price.

Due to its importance, measuring CLV has been one of the most important tasks for scholars and practitioners in the field of marketing. Accordingly, various methods have been developed to predict and measure CLV in various contexts and industries settings, including calculating CLV for marketing resource allocation in the airline industry setting (Rust et al., 2004), business-to-business setting (Venkatesan and Kumar, 2004; Kumar et al. 2008), and online retail market (Dahana et al., 2019). These methods are generally based on key assumptions concerning retention rate and profit margin while incorporating the cost of acquisition, retention, and service. While this approach has been widely used and has proved to be useful in many settings, it has several limitations. One major limitation of CLV is that it assumes a company does not react to changes in the market once it has invested in a customer, which is not realistic (Méndez-Suárez & Crespo-Tejero, 2021). The changes in customer preferences, including churn or a reduction in product/service purchases from the company, are not factored into the CLV calculation. These changes represent a risk that could affect the future profitability of a customer. Therefore, to obtain an accurate estimation of customers' value, valuation techniques must account not only for customer profitability but also for associated risks (Buhl & Heinrich, 2008; Yun & Yan, 2013). Thus, it is crucial to account for risk factors when assessing customers' value, as failing to incorporate these parameters in the CLV calculation could lead to incorrect customers' value estimations.

1.2 Problem Context

Despite the importance of accurately estimating CLV, studies that take into account the impact of risk on CLV calculation are limited (Singh & Singh, 2016). One of the pioneering works in this area was presented by Dhar and Glazer (2003), who argued that companies often fail to consider whether all their valuable customers are collectively desirable from a risk perspective. To address this gap, Dhar and Glazer (2003) proposed a new metric called risk-adjusted lifetime value (RALTV), which takes into account the impact of risk on customers' value. While several studies have since emerged that apply customers' risk to CLV calculation in different industries and contexts, the majority of these studies have been conducted in the financial services industry (FSI), including insurance, banking, and peer-to-peer lending. Moreover, the main application of these studies has been to determine the optimal customer portfolio composition using mean-variance methodology, with the volatility of customer income serving as the primary source of risk (e.g., Buhl & Heinrich, 2008; Homburg et al., 2009; Tarasi et al., 2011; Juhl & Christensen, 2013).

Furthermore, several other applications of the study exist, such as calculating more accurate CLV to find the most profitable customers for marketing purposes (Singh et al., 2013; Machado & Karray, 2022a) or using risk-adjusted customers' value for adaptive pricing in e-commerce (Ruch & Sackmann, 2012). These potential applications demonstrate the importance of incorporating customers' risk into customers' value calculations. Given that customers' risk exists in all industries, it is imperative to incorporate the impact of risk on the calculation of customers' value to obtain accurate estimations. Moreover, the nature of customers' risk may vary across industries, meaning that a standard approach cannot be applied across the board. For example, the probability of default (PD) risk in financial services cannot be used in other industries. Hence, it is essential to conduct more in-depth research to explore the impact of customers' risk on CLV calculations and expand the scope of existing studies to ensure accurate estimations of customers' value across various industries.

Furthermore, advances in technology and the rise of big data have facilitated the collection and analysis of extensive customer data. As a result, customer relationship management (CRM) has undergone a significant transformation, with companies now benefiting from more precise and actionable insights into customer behavior. Moreover, the emergence of machine learning (ML) has introduced new options for predicting the impact of customers' risk on value. ML can analyze large amounts of customer data more efficiently and accurately than traditional methods, which could lead to more precise predictions of future customers' value (Borle et al., 2008). In the quest to optimize customer portfolios, it is essential to adapt and extend existing models to incorporate data-driven approaches that enable more precise predictions of future customers' value. However, despite its growing popularity, only two studies fully utilize ML to predict risk-adjusted CLV. Both studies are conducted by Machado & Karray (2022a, 2022b) in an FSI setting. Therefore, further research is needed to explore the potential of ML in this area. A comparative analysis between ML and traditional methods for calculating risk-adjusted customers' value could provide valuable insights into their effectiveness and potential applications across various industries. More accurate and precise CLV calculations can help organizations make informed decisions regarding customer acquisition and retention strategies (Gupta et al., 2004).

1.3 Research Objective and Research Question

Although several studies address the importance of incorporating customers' risk into CLV calculations, there remains a lack of research focused specifically on the Business-to-Consumer (B2C) industry, particularly within the non-contractual customers of the telecommunication sector. This study addresses the current research gap by developing a risk-adjusted CLV metric tailored specifically to the B2C industry, particularly for non-contractual customers within the telecommunication industry. Furthermore, while some studies use ML algorithms to predict risk-adjusted CLV, their application to the telecommunication industry remains largely unexplored. The research objectives are two-fold: firstly, to introduce a new metric that considers the impact of risk on customers' value, with a specific focus on telecommunication industry applications. Secondly, to explore the potential of ML algorithms in predicting the risk-adjusted CLV, to improve the accuracy and precision of CLV calculations. By achieving these objectives, this study aims to provide a more comprehensive understanding of customers' value and help telecommunication companies make more informed decisions regarding customer acquisition and retention strategies.

To accomplish the research goal, this study attempts to address the following research question (RQ):

How can machine learning techniques be effectively applied to predict risk-adjusted customer lifetime value in the non-contractual (B2C) setting of the telecommunication industry?

The research consists of various sub-questions (SQ) aimed at gaining a comprehensive understanding of the existing literature. The initial six questions focus on establishing a foundation for the study.

- SQ1. How has the incorporation of customers' risk into CLV calculation evolved over time in the literature?
- SQ2. What are the industries or domains where customers' risk has been incorporated into CLV calculation?
- SQ3. What are the commonly used methods for incorporating customers' risk into the calculation of customers' value in the industry?
- SQ4. What is the state-of-the-art ML model used to predict the risk-adjusted CLV in the Telecommunication industry?
- SQ5. What is the most significant customer's type of risks to be considered when assessing customers' value in the Telecommunication industry?

SQ6. How much historical data and what time period should be considered to evaluate the risk-adjusted customers' value in the industry?

Using the insights gained from the literature review, a research design is developed to build a ML model to predict the risk-adjusted CLV in the telecommunication industry. This model address the following sub-question:

SQ7. What specific risks identified from the literature are most relevant to the telecommunication industry, and how can these be quantified for inclusion in the ML model?

SQ8. How to develop a ML Model to predict the risk-adjusted CLV in telecommunication industry?

Finally, the ML model is tested and evaluated. Furthermore, the model's performance and validity are assessed by evaluating the results using different data splitting strategies. Additionally, the results of feature importance are compared with the traditional results. In doing so, this part attempts to answer the following sub-question:

SQ9. How does the strategy of splitting data into training and test sets affect the performance of the ML models for risk-adjusted CLV prediction?

SQ10. What is the most important feature/variable to predict the risk-adjusted CLV in telecommunication industry?

SQ11. How does the most important feature from the model compared to the traditional calculations method?

1.4 Research Scope and Limitations

The scope of this research involves identifying customers' risk and proposing a novel method to incorporate this risk into the CLV calculation, thereby creating a RAR metric. Subsequent steps include examining the statistical significance of the proposed approach to confirm its distinctiveness. Once the uniqueness of the RAR is established, an ML model is developed to predict the RAR, and the model's performance is evaluated. To validate the model's accuracy, a feature importance analysis is conducted. This analysis facilitates a comparison between the most influential features in the model and the variables used in traditional CLV calculation methods.

The study is conducted in collaboration with one of the largest mobile telecommunications providers in Indonesia, and the data used are from prepaid customer data. The sample is randomly drawn from all geographic areas in Indonesia, as the company operates nationwide. Customers' risk in this study is related to the telecommunications industry, such as customer churn and income volatility. The dataset consists of a total of 200,000 customers, and the data observed span a period of 3 years, from January 2020 to December 2022.

Regarding limitations, this research may have certain constraints. Firstly, the findings and conclusions drawn from the study may only be applicable to the specific context of the telecommunication industry in Indonesia and may not be generalizable to other industries or regions. Additionally, the accuracy of the CLV predictions using ML algorithms may depend on the quality and availability of data, as well as the performance of the ML model used. The study acknowledges that there may be limitations in the available data or potential issues with the model that could impact the results. Moreover, the study may not be able to capture all possible customers' risks, and there may be other factors that could impact CLV but are not considered in this research (e.g., the probability of customers switching segments). Lastly, as the study is based on data from a single mobile telecommunications provider, the findings may not necessarily represent the entire industry or all types of customers. Despite these limitations, the research aims to contribute to the understanding of risk-adjusted CLV in the telecommunication industry and provide insights for future research in this area.

1.5 Research Methodology

This research adheres to the Design Science Research Methodology (DSRM) defined by (Peffers et al., 2007). DSRM consists of six steps arranged in a sequential order, which include problem identification and motivation, defining solution objectives, design and development, demonstration, evaluation, and communication. Figure 1 illustrates these six steps.

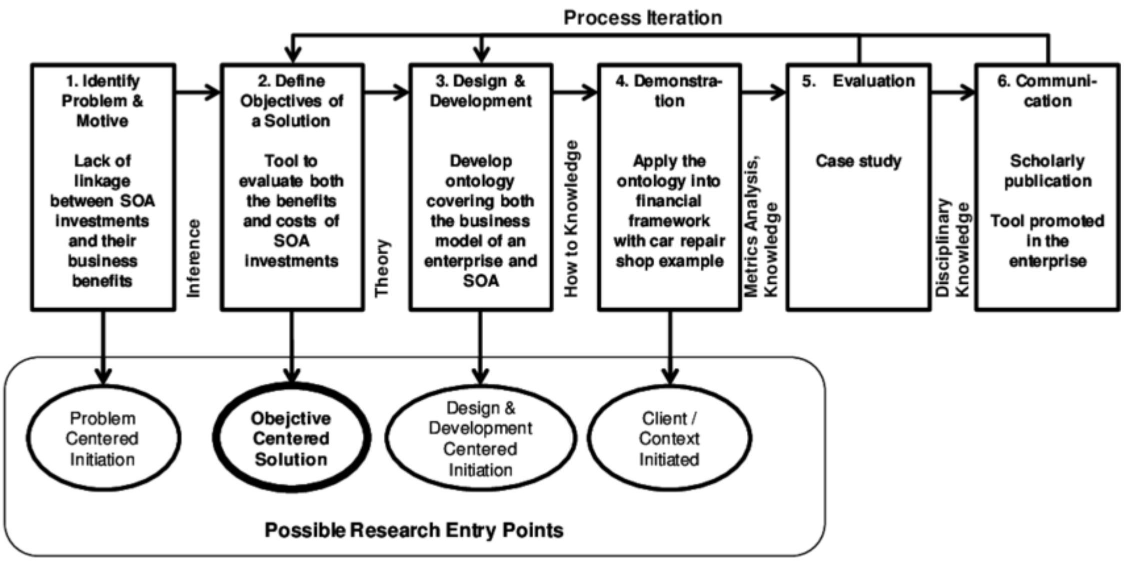


Figure 1 Design science research methodology (Peffers et al., 2007)

1. **Problem Identification and Motivation**: The first step of the DSRM process is to identify the problem and justify it with motivation. The thesis provides a clear overview of the problem identification and motivation, which can be found in Chapter 1.
2. **Define the objective for a solution**: The second step in the DSRM process is to define the objectives for the solution based on the identified problem and assess its feasibility. The objectives of this research are stated in Sections 1.3 and 1.4. To achieve these objectives, a systematic literature review was conducted in Chapter 2 to gather all relevant information for the research question and develop a metric for predicting risk-adjusted CLV using ML algorithms. Chapter 2 provides detailed responses to all SQ1-SQ6, thoroughly reviewing the available literature.
3. **Design and development**: The third phase of the DSRM process involves several activities, including the extraction and analysis of data, the identification of customers' risks from the data, and the development of a metric for risk-adjusted CLV. The specific activities involved in this phase are presented in more detail in Chapter 4.
4. **Demonstration**: The next step is to demonstrate the practical application of the developed solution. This can take the form of experiments, simulations, or case studies. For this study, the developed ML model to predict risk-adjusted CLV is presented and discussed in Chapter 4 and Chapter 5.

5. **Evaluation**: At this stage, the results of the ML model are presented and evaluated. The evaluation includes identifying the most important features for predicting risk-adjusted CLV using Explainable AI (XAI) and comparing the predicted values with those obtained using traditional CLV methods. A detailed account of the evaluation process and findings can be found in Chapter 5.
6. **Communication**: The final stage of the research aims to disseminate the findings and outcomes of the study. This is achieved through a comprehensive report that presents the research process, challenges, and artifacts, as well as novel insights and relevant information that can help stakeholders understand the research problem and its solutions. The report targets organizations, researchers, and audiences who are interested in the topic and provides recommendations for future research and practical applications.

1.6 Thesis Outline

This thesis is systematically organized into six chapters to provide a cohesive understanding of the effective application of ML techniques in predicting risk-adjusted CLV in the non-contractual, B2C context within the telecommunication industry.

Chapter 1 serves as the foundation, introducing the research topic, objectives, the central research question, and the array of sub-questions that guide this research. This chapter sets the stage for the following chapters by providing an overview of the current situation and outlining the research scope.

Chapter 2 is dedicated to a comprehensive literature review that aims to address sub-questions SQ1 through SQ6. The exploration of past and current academic literature paves the way to understanding the evolution and applications of risk-adjusted CLV across various industries and domains. It further delves into identifying prevalent methodologies employed within the industries. Crucially, this chapter identifies and discusses the gap in existing literature that this study aims to fill.

Chapter 3 outlines the research methodology employed in this study. It covers the philosophical and analytical underpinnings of the chosen methodological approach, providing justification for the selected methods.

In Chapter 4, the proposed RAR model is presented. The chapter elaborates on the formula employed to calculate RAR, highlighting any assumptions made. Further, the selected dataset and its characteristics are described, along with the construction of the ML model used in the study, which includes data preprocessing, feature engineering, model development, and the model evaluation and validation process.

Chapter 5 presents the results derived from the application of the RAR model and ML techniques. This chapter provides a detailed analysis of the calculated RAR values, the performance of the ML model, and the results of the model validation process. It also offers an insightful discussion on the importance of different features, their influence on model predictions, and a comparative analysis with traditional calculation methods.

Finally, Chapter 6 concludes the thesis by addressing the main and sub-research questions, discussing the implications and limitations of the study, and providing suggestions for future research.

2 Literature Review

This chapter presents the background information and a systematic literature review (SLR) conducted to answer the research questions outlined in Chapter 1. The background information presents the background and definitions of useful concepts and terms concerning ML and its applications to establish a contextual understanding for the audience. Whereas the SLR aims to investigate the historical development of studies that have incorporated customers' risk into CLV calculation (SQ1), as well as the industries or domains in which these studies have been applied (SQ2). Further, the SLR also uncover the general methods used to incorporate customers' risk into CLV calculation in the industry (SQ3) and identify the ML methods used in previous studies (SQ4). The review also examines the types of customers' risks considered (SQ5) and the total customer data and time period of observation used to evaluate customers' risk in the telecommunication industry (SQ6). Finally, the literature review aims to identify the gaps in the existing literature and provide insights for the development of a ML model to predict risk-adjusted CLV in the telecommunication industry.

2.1 Background Information

2.1.1 Machine Learning

ML is a critical subfield of artificial intelligence that uses algorithms and statistical models to teach computers how to learn from data and make decisions or predictions without explicit instructions (El Naqa & Murphy, 2015). It proves to be incredibly powerful, surpassing human performance in various domains (Weld & Bansal, 2019). As the amount of data in various fields continues to grow exponentially, ML becomes an indispensable tool for analyzing and making sense of this data (Mahesh, 2018). Its applications are incredibly diverse, ranging from pattern recognition, computer vision, and spacecraft engineering to finance, entertainment, ecology, computational biology, and biomedical and medical fields (El Naqa & Murphy, 2015). There are three main types of ML depending on the availability of feedback to support the learning process: supervised learning, unsupervised learning, and reinforcement learning.

Supervised ML models are designed to develop a function that accurately predicts labels for unseen data, using labelled examples as a training set. The algorithm undergoes a training process that enables it to identify patterns and learn from observations, thus allowing for insightful predictions. This iterative process continues until the algorithm reaches a high degree of accuracy and performance. Typical tasks, such as classification and regression, use supervised ML algorithms because the desired output is already known (Alpaydin, 2010). Applications of supervised learning in the marketing field include predicting customer churn (Al-Mashraie et al., 2020) and estimating customers' value (Tsai et al., 2013).

Unsupervised ML algorithms, in contrast to supervised ML, are implemented in situations where labelled examples are not available, thus eliminating the need for human instruction. These algorithms possess the innate ability to identify correlations and connections by examining the data available, employing clustering techniques to identify both similarities and disparities (Alloghani et al., 2020). Within the marketing sector, the applications of unsupervised ML are extensive, with its utility in customer segmentation (Purnomo et al., 2020) and anomaly detection (Tan et al., 2020). K-means clustering and probabilistic clustering methods are among the most commonly utilized algorithms in unsupervised learning (Alloghani et al., 2020).

Finally, reinforcement learning is a type of ML that involves learning by trial and error. Instead of being given specific labelled examples like in supervised learning, the system learns through rewards or punishments received from its environment. The goal is for the system to learn to make decisions that lead to the maximum reward. A fundamental reinforcement algorithm may be represented as a Markov Decision Process (MDP), which specifies a set of states, actions, rewards, and transition probabilities that account for specific times, actions, and states (Alpaydin, 2010). An example of reinforcement learning application in the marketing field is dynamic pricing (Liu et al., 2021).

2.1.2 Explainable AI/ML (XAI)

XAI is a field of research that aims to create ML models that are transparent and explainable to humans. The term XAI was first introduced by Van Lent et al. in 2004 to describe the ability of their system to explain the behaviours of AI-controlled entities in simulation games applications (Adadi & Berrada, 2018). However, the idea of XAI can be traced back to the mid-1970s and early 1980s, when some expert systems explained their results via the applied rules (Xu et al., 2019). The need for XAI arises due to the increasing adoption of complex ML models such as Deep Neural Networks (DNNs). These models are often referred to as "black box" models because their internal decision-making processes are not transparent to humans, making it difficult for humans to understand their decision-making process. In fact, the output of DNNs cannot be explained by the network itself, nor by an external explanatory component, and not even by the developer of the system, which makes them particularly challenging to explain (Xu et al., 2019). This lack of transparency is particularly problematic in critical domains such as healthcare, finance, and justice, where the decisions made by these models can have significant real-world impact (Samek & Müller, 2019). As a result, there is a growing demand for ML models that can provide explanations or justifications for their predictions. This would enable humans to understand the factors that influenced the decision and increase trust in the model.

In addition to increasing trust in ML models, XAI has several other benefits. One of the main benefits is that it can help identify and correct biases in the data or the model itself, which can have significant real-world implications (So C. , 2020; Doshi-Velez & Kim, 2017). XAI also has the potential to improve the overall performance and effectiveness of ML models. By providing explanations for the model's predictions, developers and domain experts can gain insights into the model's decision-making process and identify potential areas for improvement or refinement (Guidotti et al., 2018; Xu et al., 2019). Finally, XAI can aid in compliance with ethical and legal regulations by providing justifications and explanations for the decisions made by the model.

In recent years, there has been a growing interest among AI researchers to create transparent systems by opening the "black-box" of neural networks. The field of XAI can be categorized into two main strands of work: transparency design and post-hoc explanation (Lipton, 2018; Linardatos et al., 2021), as illustrated in Figure 2. The transparency design strand aims to provide developers with an understanding of how a model functions by focusing on the model structure, such as the construction of a decision tree, individual components, such as a parameter in logistic regression, and the training algorithms, such as solution seeking in a convex optimization. Post-hoc explanation techniques, on the other hand, focus on explaining why a particular result is inferred, from the perspective of users. These techniques include giving analytic statements, providing visualizations, and giving explanations by example. SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016) are notable examples of post-hoc explanation methods widely used in XAI research. These techniques provide insights into the contribution of individual features or components in a model's predictions, allowing for more interpretability and transparency in AI systems.

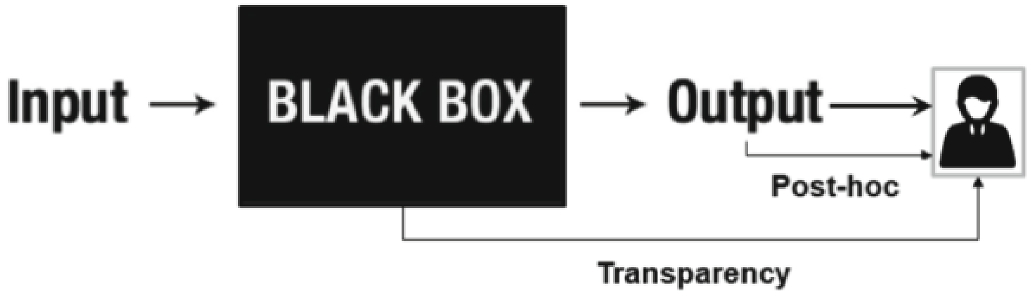


Figure 2 Two categories of Explainable AI work: transparency design and post-hoc (Xu et al., 2019)

XAI has garnered considerable attention worldwide, not only in the realm of research, but also in industry. In April 2017, the United States Defence Advanced Research Projects Agency (DARPA) [1]initiated the XAI program with the goal of enhancing the explainability of AI decisions while maintain the model performance. Similarly, in July of the same year, the Chinese government issued "The Development Plan for New Generation of Artificial Intelligence," aimed at promoting high-explainability AI and strong-extensibility AI. Furthermore, in May 2018, the European Union published the "General Data Protection Regulation" (GDPR)[2], which grants citizens a "right to explanation" when they are affected by algorithmic decision-making (Xu et al., 2019). As such, XAI has become increasingly crucial for all stakeholders, including users, those affected by AI decisions, and developers of AI systems.

2.1.3 Machine Learning and Explainable AI in CLV Predictive Frameworks

The application of ML for CLV prediction has gained attention in comparison to traditional methods, as it shown to outperform traditional models (Borle et al., 2008). The earliest study that applied ML algorithms in customer valuation is the study of (Gelbrich & Nakhaeizadeh, 2000) where they use linear regression and multiple linear regression to predict the purchase frequency, price acceptance and the discount rate in the automotive industry. In 2013, Tsai et al. presented a novel approach for predicting CLV using a hybrid ML model. Their model integrates the clustering algorithm K-means and three classification algorithms, namely Decision Tree, Logistic Regression, and Multilayer Perceptron. The authors' findings

[1] https://www.darpa.mil/
[2] https://gdpr-info.eu/

suggest that the hybrid model exhibits superior performance in the domain of CLV prediction compared to using a single model. Furthermore, the study highlights the criticality of selecting appropriate clustering and classification techniques when constructing a hybrid model. In recent research conducted by Asadi and Kazerooni (2023), it was demonstrated that stacked ensemble learning outperformed several popular predictive methods such as deep neural networks, bagging support vector regression, light gradient boosting machine, random forest and extreme gradient boosting in predicting CLV. These findings suggest that there have been numerous studies conducted on predicting CLV using ML, and that more advanced ML algorithms are being explored to improve accuracy and performance.

Despite numerous studies on ML-based CLV prediction, the use of XAI in customer valuation is limited. To the best of the author's knowledge, only one study by Yılmaz Benk et al. (2022) that incorporate XAI in their customer valuation study. They has predicted customer valuation in e-commerce using CLV and two other metrics: distinct product category and trend in amount spent. They proposed using a multi-output deep neural network (DNN) to identify the most profitable customers. In the end, Yılmaz Benk et al. (2022) derived the Shapley value using the XAI method to interpret the DNN's decisions. Although the application of XAI in customer valuation studies is limited, XAI research has been conducted in other areas related to customer valuation. For instance, XAI has been applied in churn prediction (Marín Díaz et al., 2022), customer turnover (Souza & Leung, 2021), propensity to buy a product (Gramegna & Giudici, 2020), and augmented cross-selling (Haag et al., 2022).

2.2 Systematic Literature Review (SLR)

SLR is a structured and transparent process that helps researchers gain a deeper understanding of prior work in their field and identify gaps and potential avenues for future research (Kraus et al., 2022). The process typically involves four phases, starting with identifying relevant databases and keywords for the search process. The selection phase follows, where inclusion and exclusion criteria are established and applied for selecting relevant studies. The third phase involves extracting and integrating data from selected studies, while the fourth phase focuses on interpreting the results to identify patterns and themes in the data (Mengist et al., 2020).

2.2.1 The Search Phase: Scientific Database and Search Queries

As the first phase in the SLR process, the search phase plays a critical role in determining the effectiveness and validity of the entire review. This phase encompasses several fundamental tasks, including identifying relevant databases for sourcing studies, and determining the keywords to be utilized in the search process.

This SLR utilizes two major academic databases, Scopus[3] and Web of Science (WoS)[4], as primary sources of information. These databases were selected due to their ability to provide comprehensive coverage of academic literature, including both recent and historical publications, relevant to this topic (Harzing & Alakangas, 2016). Scopus and WoS are widely regarded as high-quality sources for academic research because they employ rigorous evaluation processes to ensure that only high-quality publications are included in their databases (Baas et al., 2020; Pranckute, 2021). Furthermore, both databases are indexed and abstracted, meaning that they provide bibliographic information and summaries of articles rather than the full text (Pranckute, 2021). Advanced search options offered by these databases allow users to apply selection criteria, making it easier to find relevant literature. Overall, Scopus and WoS are appropriate sources for this SLR study due to their comprehensive coverage and rigorous evaluation processes.

Further, the search query applied in the database is formulated based on a set of keywords related to the research questions. The main objective of this SLR is to examine the prediction of customers' value with the inclusion of customers' risk, and therefore, all variations of the name will be included in the search criteria. However, abbreviations will not be included since they may yield irrelevant results or obscure the intended meaning of the query. Furthermore, it can be assumed that papers or journal articles that use abbreviations have already been captured in the search, as the full name has been included in the query. The search strings used for all sources comprises of strings as follows:

1. ("Customer" AND ("Risk Adjusted Revenue" OR "Risk adjusted lifetime value" OR "Risk-adjusted Lifetime value"))
2. ("Customer Lifetime Value" OR "Customer Value" OR "Customer Portfolio" OR "Customer Asset") AND ("Optimization" OR "Risk-Adjusted" OR "Risk Adjusted")

Both search string was combined to get all result combination. In order to further control the relevance of the search result, the search query is applied in the article's title, abstract, and keywords.

---

[3] https://www.scopus.com
[4] https://webofknowledge.com

2.2.2 The Selection Phase

To ensure the quality and reliability of the research, only peer-reviewed articles published in reputable journals or conference proceedings were included. The articles adhered to proper formatting guidelines, including complete author identification and publication information. There was no restriction on the publication year in order to provide a comprehensive overview of the topic. Only articles published in English were included in the study, while non-English articles were excluded. In addition, articles that did not have a clear connection to the research questions of this study based on their title, abstract, or content were also excluded. Duplicate articles with identical titles or content found in multiple databases were also excluded. Finally, incomplete or excessively short articles were not considered for the study. Table 1 provides a summary of the inclusion and exclusion criteria adopted by this study during the SLR.

Table 1 Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
Studies published in conferences proceeding and journal articles	Articles that are not complete
Adhere to proper formatting guidelines, including complete author identification and publication information	Studies that are not related to the main RQ from title, abstract and content
English based peer reviewed studies	Duplicate articles by title or content

To ensure the relevance of the articles to this study and to avoid spending time reading irrelevant publications, the gathered articles underwent a thorough review process. This process involved multiple steps, beginning with executing the defined search queries on each scientific database, followed by applying the inclusion and exclusion criteria as described in Table 1. The duplicate studies from different databases were removed, and the irrelevant studies were excluded based on the assessment of their title and abstract. Afterwards, the remaining studies were evaluated through the analysis of the full text and exclusion of short, overly general, or incomplete studies. Finally, the primary papers that were used in this research were selected.

At the conclusion of this review process, 22 articles were selected from a pool of 386 articles identified in the initial step. Figure 3 depicts the SLR selection flow chart used in this study.

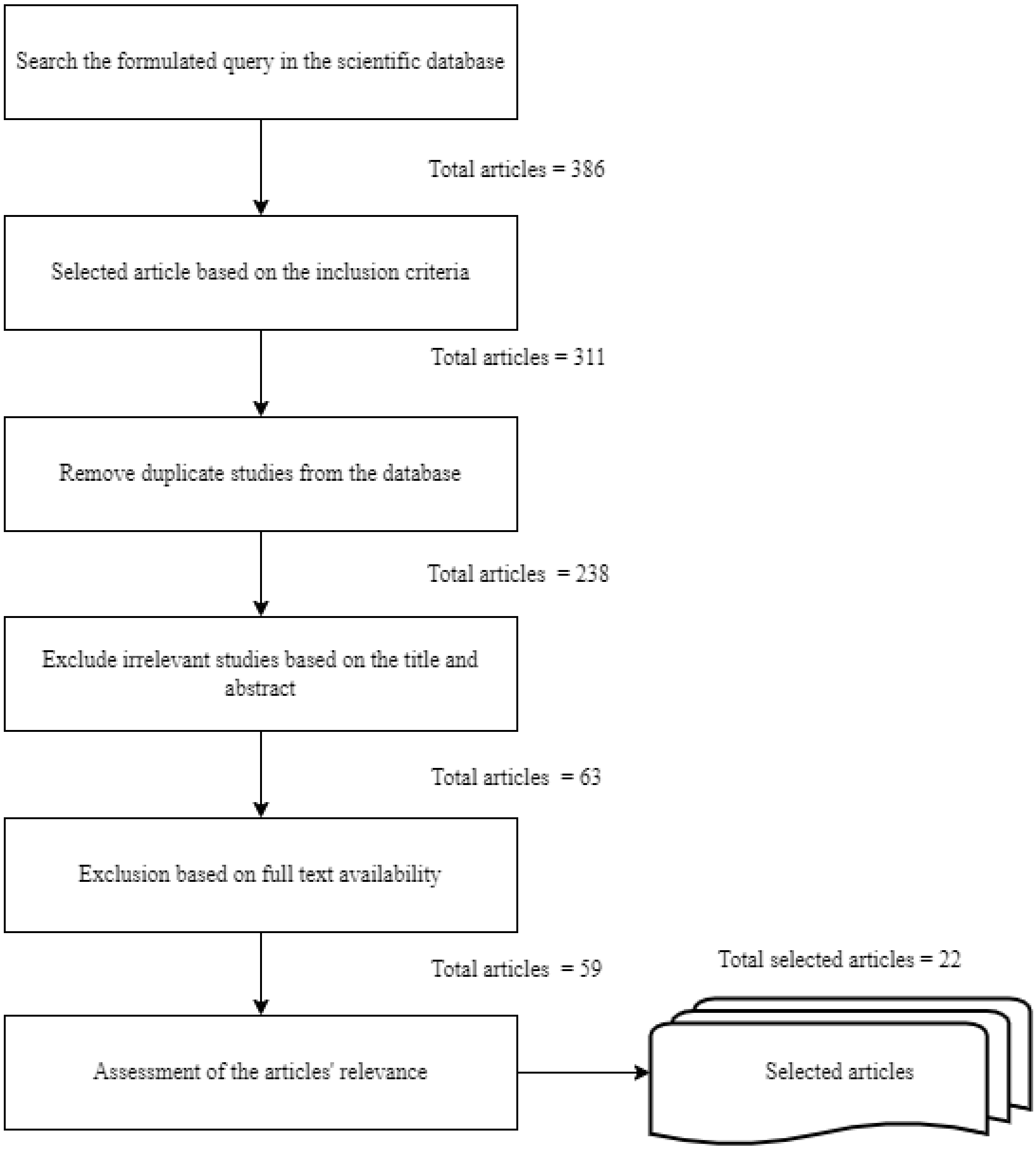


Figure 3 SLR selection flow chart

2.3 SLR Results

This section provides an overview of the research findings from the SLR. Subsection 2.3.1 presents details about the research method and output of the papers found in the SLR, including co-citation analysis and general themes discussed in the study. In the subsequent Subsections (2.3.2 – 2.3.7), each of the SQ1-SQ6 identified in the study will be answered. Through analysis of the SLR findings, this study aims to offer a comprehensive overview of the current state of research on the topic and identify gaps in the existing literature that need to be addressed.

2.3.1.1 Data Extraction and Synthesis

After selecting the relevant articles, the SLR study moved forward to the third phase, the data extraction and synthesis, aimed at extracting information pertinent to addressing the research questions. This involved a thorough reading of all studies while utilizing a research framework designed to answer the research questions. Table 2 presents the 22 selected articles along with the research method used and the study's output. Research methods included Literature Review (LR), Experiment (E), and Comparative Study/Validation Study (CS/VS), each of which could produce varying output categories (Theoretical (T), Conceptual Model (CM), and Empirical results (ER)). The majority of the literature provided theoretical perspectives on the subject matter. These studies were classified as theoretical (T) when they presented overarching concepts, advantages, disadvantages, or design principles for incorporating risk when calculating customers' value. Additionally, some studies offered a conceptual model (CM) that supplemented the theoretical framework by providing graphical representations of the design decisions inherent in the model. All of the studies also produced empirical results (ER) derived from calculating (C), predicting (P), or both (CP) customers' value while incorporating risk. Similarly, this study also produced ER derived from both calculating and predicting the risk-adjusted revenue.

Table 2 Research method and output

Reference	**Research Method(s)**			**Output**		
	LR	**E**	**CS/VS**	**T**	**CM**	**ER**
Dhar & Glazer (2003)			v	v		C
Ryals (2003)			v	v		C
Ryals & Knox (2005)	v		v		v	C
Wangenheim & Lentz (2005)	v		v	v		CP
Hai-wei et al. (2006)		v				P
Ryals & Knox (2007)	v		v	v	v	C
Buhl & Heinrich (2008)	v		v	v		C
Homburg et al. (2009)	v		v	v		CP
Sackmann et al. (2010)	v		v	v		C
Tarasi et al. (2011)	v		v	v		C

Ruch & Sackmann (2012)			v	v		C
Singh et al. (2013)	v		v	v		C
Juhl & Christensen (2013)	v		v	v		C
Albadvi & Norouzi (2013)			v			CP
Yun & Yan (2013)			v			C
So et al (2014)			v			P
Petersen & Kumar (2015)		v		v		C
Norouzi & Albadvii (2016)	v		v	v		CP
Singh & Singh (2016)	v		v	v		C
Viviani, Komura, & Suzuki (2021)	v		v	v	v	P
Machado & Karray (2022a)			v			P
Machado & Karray (2022b)	v		v		v	CP
This Study (2023)	**v**	**v**	**v**	**v**		**CP**

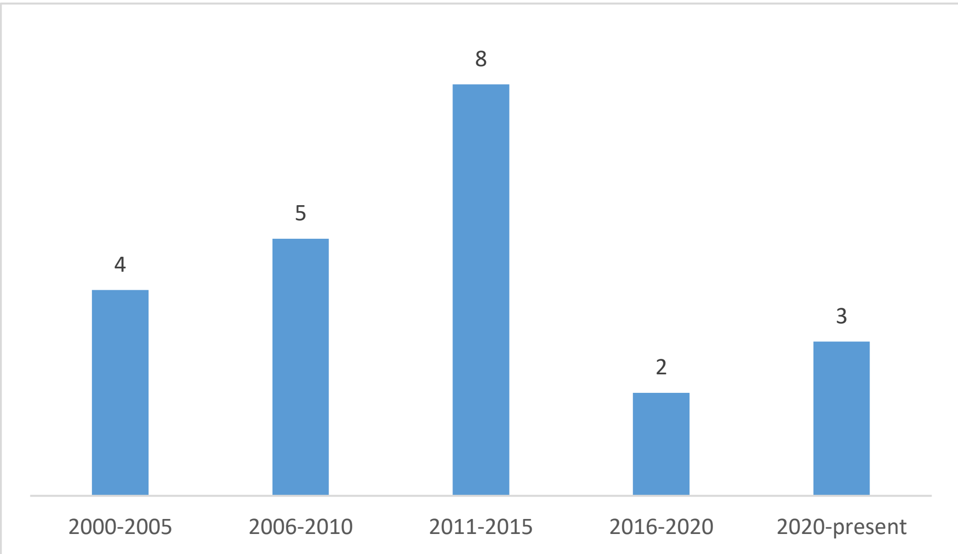


Figure 4 Number of studies by year of publication

The results of SLR method in this study found articles between 2003 to 2022 (Figure 4). The earliest study identified was conducted by Dhar and Glazer in 2003. On average, one to two papers were published annually on the topic, with four studies found in 2013. A five-year gap in publications occurred after 2016, with one study identified in 2021 and two additional studies produced by Machado and Karray in 2022 (A detailed overview of publication per year is presented in the Appendix 1). The majority of the studies were published in marketing and management journals, with 41% in marketing journals (e.g., Journal of Marketing, Journal of Marketing Research) and 36% in management journals (e.g., Management Research Review). The remaining studies were published in other journals, such as the European Journal of Operational Research and Decision Support Systems journal. In terms of number of citations, the early studies conducted between 2003 and 2011 received a high number of citations, exceeding 100 in most cases. Citations are commonly used as a measure of a publication's usefulness, impact, or influence. The number of citations received by a publication is often taken as an indicator of its influence (Aksnes et al., 2019). Table 3 presents a detailed breakdown of the citation distribution.

Table 3 Number of citations

Number of citations	Number of studies	List of studies
0-50	14	Wangenheim & Lentz (2005); Hai-wei et al. (2006); Sackmann et al. (2010); Ruch & Sackmann (2012); Juhl & Christensen (2013); Norouzi & Albadvi (2013); Singh et al. (2013); Yun & Yan (2013); So et al (2014); Norouzi & Albadvi (2016); Singh & Singh (2016); Viviani, Komura, & Suzuki (2021); (Machado & Karray (2022a, 2022b)
50-100	1	Buhl & Heinrich (2008)
100-150	2	Tarasi et al. (2011); Ryals & Knox (2005)
150-200	3	Ryals (2003); Petersen & Kumar (2015); Homburg et al. (2009)
>200	2	Dhar & Glazer (2003); Ryals & Knox (2007)

A citation network analysis within the SLR result was conducted and is presented in Figure 5 to further examine the relationships between the studies. In this figure, the circle size depicts the number of papers that cite the respective paper in the network. Additionally, the arrow lines indicate network correlation: the paper at the head of the arrow is the one that's been cited, while the paper at the tail of the arrow is the citing paper. The analysis reveals that Dhar & Glazer's (2003) work is the most influential study among the SRL results, with 17 out of the 21 studies included in the SLR citing them. Tarasi et al. (2011) and Buhl & Heinrich (2008) also hold significant importance and have been cited by 9 and 8 studies, respectively. Co-citation analysis shows that these three studies are often cited together, with 8 out of 9 studies that cite Tarasi et al. (2011) and 7 out of 8 studies that cited Buhl & Heinrich (2008) also citing Dhar & Glazer (2003). This indicates a strong relationship and research focus among the three most important studies (Annarelli et al., 2021). Five studies did not cite or were not cited by the other studies, most of which were published after 2011. Two studies, Hai-wei et al. (2006) and Petersen & Kumar (2015), were found to not cite and citing the other studies found in SLR.

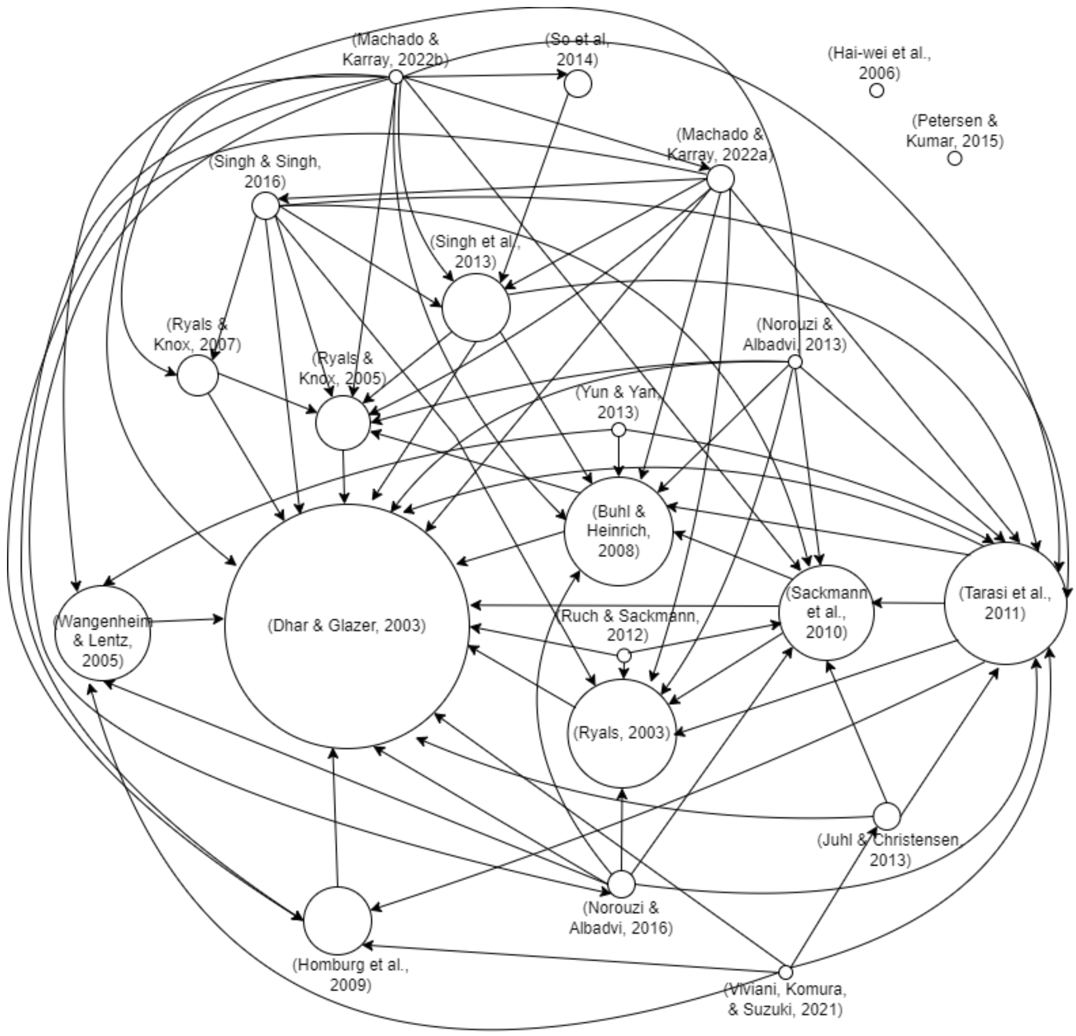


Figure 5 Citation network

The most dominant theme is evaluated with the analysis of word clouds. Figure 6 (A) demonstrate the high-frequency word in the titles of research papers, where the word size represents the frequency used in the titles. The most popular words in the title of selected papers are “customer”, “risk”, “portfolio”, “value” and “model”. In Figure 6 (B), the most popular keywords of the selected studies are “customer relationship management”, “customer portfolio management”, “risk”, “customer lifetime value” and “customer portfolio”. These results suggest that the majority of the studies focus on the relationship between "risk" and "customer," with a particular emphasis on "customer relationship management," "customer portfolio management," and "customer lifetime value".

(A) (B)

Figure 6 Word cloud of (A) titles and (B) keywords of selected studies

The studies mentioned in Table 2 were further explored, and information was extracted to find the relevant evidence to answer the research question. The detailed results of the data extraction are shown in Table 4, which will be structured with the following elements:

- Reference;
- Research theme (the key aim of the study);
- Industry setting of the study;
- Type of risk used in customer lifetime value calculation;
- Methodology used in the study;
- Time frame of the data; and
- Total observed customer/data.

The following section of this chapter presents the fourth phase of the SLR study. It further reviews the studies found and discusses the temporal evolution of studies, the application area, the risk used in the calculation, and the methodology used to incorporate risk in customers' value calculation.

Table 4 Data extraction and synthesis of prior studies

Reference	**Research Theme**	**Industry**	**Risk**	**Method**	**Time frame (years)**	**Total Customer**
Dhar & Glazer (2003)	Find the best customers' segment by calculating Risk Adjusted Lifetime Value (RALTV)	• B2B	• Beta risk	• Mean Variance	• 5	• -
Ryals (2003)	Introducing risk in CLV calculation	• FSI (Insurance)	• Churn	• Mathematical and statistical approach	• 4	• 2
Ryals & Knox (2005)	Introducing risk into CLV calculations within the insurance industry involves forecasting revenue, risk, and costs, and then calculating the risk-adjusted CLV.	• FSI (Insurance)	• Insurance claim risk • Churn	• Mathematical and statistical approach • Forecasting	• 4	• 12
Wangenheim & Lentz (2005)	Attempt to segment the customers based on metrics that describe the dynamics of their lifecycle development, including return and risk characteristics.	• B2C (Airlines)	• Beta risk • Variance of customer transaction • Inactive period	• Regression • correlation • clustering (Kmeans)	• 4	• 23652
Hai-wei et al. (2006)	Incorporating risk into the measurement of customers' value through a Bivariate Hierarchical Bayesian approach.	• B2B	• Churn • Decline of purchase • Volatility of income	• NBD • Mathematical and statistical approach	• 2	• 2343
Ryals & Knox (2007)	Adjusting the CLV calculation by incorporating relationship risk.	• FSI (Insurance)	• Relationship risk (customer	• Mathematical and statistical approach	• 4	• 10

	Developing a customer relationship scorecard as a managerial tool to assess risks in customer relationships and formulate risk mitigation strategies.		overall relationship, account relationship, knowledge of customer)			
Buhl & Heinrich (2008)	Balancing the customer portfolio across various customers' segments based on their occupations, estimating the return, risk, and correlation coefficients of different groups according to their annual revenue. This balancing is achieved through statistical analysis (mean-variance) and consideration of correlations between segments.	• FSI	• Beta risk	• Mean Variance	• 10	• -
Homburg et al. (2009)	Developing a CLV model for dynamic customer portfolio analysis that considers switching patterns between customers' segments over time.	• FSI • B2C (Telecommunication, Pharmaceutic, Chemical, Bank)	• Volatility of income (probability of changing segment over time)	• Regression • Mean Variance	• 1.25 (telco) • 4 (others)	• 300000 (telco) • 3422 (pharmacy) • 826 (chemical) • 100000 (bank)
Sackmann et al., (2010)	Finding the optimal mix of customers' segments (relationship-oriented and transaction-oriented) within a non-	• B2C (Commerce)	• Volatility of Income	• Mean Variance	• 0.75	• 2357

	contractual customer portfolio using Mean-Variance (MV) optimization.					
Tarasi et al. (2011)	Segment the company based on cash flow and utilize Mean-Variance analysis to determine the most profitable portfolio considering the risk-return ratio.	• B2B	• beta risk	• Mean Variance	• 7	• 516
Ruch & Sackmann (2012)	Calculate the risk-adjusted price for customers in an e-tailer. The price is adjusted based on customers' value, turnover risk, and payment risk.	• B2C (Commerce)	• Payment risk • Churn risk	• NBD • Mathematical and statistical approach	-	-
Singh et al. (2013)	Calculating Risk Adjusted Revenue in the credit card industry using DEA, and then ranking customers from the most to the least profitable.	• FSI (Credit card)	• Beta risk • Volatility of income • Probability of Default	• Data Envelopment Analysis (DEA)	• 3	• 1700
Juhl & Christensen (2013)	Building upon the work of Tarasi et al. (2011), this study aims to enhance it by incorporating growth rate considerations. It involves segmenting customers based on geographical and customers' segments, utilizing an expanded dataset, and introducing additional metrics such as reward-to-volatility (RtVo) and Jensen's α.	• B2B	• Beta risk	• Mean Variance	• 4	• 968
Albadvi & Norouzi (2013)	Calculating risk-adjusted CLV and comparing it with the traditional CLV. The process involves clustering	• FSI	• Beta risk • Downside beta	• NBD	• 2.75	• 3632

	customers based on RFM variables, estimating beta and downside beta for each segment, and subsequently calculating a risk-adjusted discount rate for each segment. The CLV is computed using the Pareto/NBD model.			• Mathematical and statistical approach		
Yun & Yan (2013)	Developing a customer portfolio optimization model under risk conditions, utilizing multi-phase marketing resource inputs as decision variables. This model is solved using a particle swarm optimization (PSO) method, aiming to assist companies in optimizing customer equity and enhancing their marketing strategies.	• B2B (Commodities)	• Volatility of income	• Mathematical and statistical approach	• 1	• 20
So et al (2014)	Employing credit card users as either transactors or revolvers, this study demonstrates the process of constructing a scorecard to predict the likelihood of each applicant being a transactor. The impact of this distinction on profitability modelling is then showcased.	• FSI	• Probability of Default	• Statistical technique (scorecard predictive model)	• 4	• 6000
Petersen & Kumar, (2015)	Conduct research and experimentation to develop a novel CLV metric that incorporates product returns as a risk factor in the calculation. Compare the results with the company's existing strategy and a benchmark model. Utilize	• B2B (Commerce)	• Risk of product return	• Mathematical and statistical approach	• 3	• 26000

	the Pareto/Non binomial distribution (NBD) model to statistically estimate the volatility of future returns generated by customers. Calculate the weight of each customers' segment and optimize the composition of customers.					
Norouzi & Albadvi (2016)	Utilizing the Pareto/Non binomial distribution (NBD) model as a statistical assumption to estimate the volatility of future returns generated by customers, calculate the weight of each customers' segment, and optimize the composition of customers.	• B2C (cosmetic)	• Volatility of income	• NBD • Mean Variance	• 2	• 1411
Singh & Singh (2016)	Develop a risk-adjusted RFM method for calculating CLV and employ DEA to determine the most profitable customer composition.	• B2B	• Churn • Volatility of Income (reaching minimum sales level & regularity of customer purchase)	• Data Envelopment Analysis (DEA)	• 2	• 2000
Viviani, Komura, & Suzuki (2021)	Enhance the profitability-risk relationship within the hotel industry by optimizing customers' segments. Employing Hidden Markov Model (HMM) for customers' segmentation,	• B2C (Hotel)	• Volatility of income from each segment	• HMM • Mathematical and statistical approach	• 5	• 5000

	the customer portfolio is subsequently optimized using a statistical approach.					
Machado & Karray (2022a)	Utilize ML to model RAR in the financial sector, integrating Customer Portfolio Theory (CPT) and Multiple Sources of Risk (MSR) approaches.	• FSI (loan provider)	• Recency of delinquency • Volatility of multiple sources of income • the credit rating. • Probability of Default	• Machine Learning	• 8	• 2 million
Machado & Karray (2022b)	Implementing a hybrid approach to evaluate Customer RAR in the financial sector. Calculate RAR using the CPT approach and group customers based on all variables in the dataset.	• FSI (P2P lending)	• Volatility of multiple sources of income (beta risk and minimum expected rate of return)	• Machine Learning	• 8	• 2 million
This Study	Introducing a novel metric to evaluate customers' RAR in the telecommunication industry. Calculate the RAR using the proposed approach and develop an ML model for RAR prediction. Subsequently, evaluate and validate the model through a feature importance analysis.	• B2C (Telecommunication industry)	• Probability of customers' churn • Beta value	• Mathematical and Statistical approach • Machine Learning	• 3	• 200000

2.3.2 Studies Modelling Risk-Adjusted CLV Throughout Time

The results of the SLR reveal that the majority of the studies that incorporate customers' risk in CLV calculations are derived from financial portfolio theory. This group of studies draws upon the work of Nobel Laureate William Sharpe, which is associated with Modern Portfolio Theory (MPT) and Capital Asset Pricing Model (CAPM) (Sharpe, 1964) and the work of Markowitz (1952), which is associated with portfolio selection theory (PST). MPT is based on the concept of share portfolios and proposes that investors aim to maximize returns for a given level of risk. The CAPM is grounded in the assumption that investors are risk averse, where investors demand higher returns for higher levels of risk. CAPM also suggest that all assets carry two distinct types of risk: systematic and unsystematic. Systematic risk is market-wide and affects all assets, while unsystematic risk is related to a single asset or a limited number of assets (Sharpe, 1964). The CAPM demonstrates that unsystematic risk can be eliminated by holding a well-diversified portfolio, whereas systematic risk cannot be diversified away (Buhl & Heinrich, 2008). Furthermore, in CAPM systematic risk is measured by beta, which represents the sensitivity of an asset's returns to movements in the overall market. Thus, in the studies that model CLV incorporating risk and are derived from the financial portfolio theory, the customer is treated as a risky asset where the risk of each customer/customers' segment is different and need to be managed to maximize the return.

Some studies, such as the studies presented by Dhar and Glazer (2003) and Ryals (2003), have applied the CAPM model to calculate a customer-specific discount rate to be incorporated into CLV models. However, Wangenheim and Lentz (2005) and Buhl & Heinrich (2008) contend that CAPM has certain drawbacks, such as disregarding unsystematic risk in the calculation. To overcome these limitations, these researchers apply the PST of Markowitz (1959) to customer portfolio management, arguing that PST model by Markowitz provides a more effective approach to mitigate these limitations. Homburg et al. (2009) have integrated customer portfolio theory and customers' segment dynamic theory, arguing that static portfolio models may overestimate the value of top-tier customers and underestimate the value of bottom-tier customers due to disregarding value dynamics in customer relationships. This approach has been further developed by Viviani et al. (2021) to determine the optimal customer composition in the hospitality industry.

Other studies have also applied the PST to customer portfolio management, such as Sackmann et al. (2010) for e-commerce, Tarasi et al. (2011) and Juhl and Christensen (2013) for B2B settings. Norouzi and Albadvi (2016) have expanded on the existing research by introducing a hybrid model that combines stochastic CLV modeling and ex-ante customer portfolio optimization, which predicts future return streams by customers and incorporates these predictions into the portfolio optimization process. Finally, Machado and Karray (2022b) propose an adaptation of the PST model by Buhl and Heinrich (2008) where they use ML to group customers and predict the RAR) value based on these groups. These studies demonstrate the value of applying financial portfolio theory in customer portfolio management.

Another approach to examine the risk-adjusted customers' value consider Multiple Source of Risk (MSR) approach, which is determined by relevant risk factors associated with the industry and business setting (Machado & Karray, 2022a). Singh et al. (2013) have observed that most studies which incorporate risk in customers' value calculations, such as those derived from the CAPM in finance, only account for one type of risk - namely, volatility of income - without identifying and quantifying different sources of risk that affect customers' value. Therefore, they proposed a framework to measure the value of credit card customers using multiple types of risk, including volatility of different sources of income and customers' PD. Subsequently, Singh and Singh (2016) employed a non-parametric approach to generate a risk-adjusted Recency Frequency and Monetary (RARFM) index for each customer, taking into account returns from various income sources and different types of risk, such as the likelihood of customer churn, the probability of reaching a minimum amount of sales, and the volatility of customer purchases. Other studies, such as Ryals and Knox (2005), have considered factors such as customers' specific insurance claims and customer churn. Finally, Machado and Karray (2022a) combine the PST and MSR approach to calculate RAR in FSI then they implement the statistical tests the evaluate the result.

2.3.3 Application Areas

The industries setting where the studies in incorporating customers' risk in the customers' value is explored. The studies found in SLR could be clustered into three big groups of industrial settings, which are FSI, B2C and B2B. The result SLR found that the majority of study was conducted in FSI (44%) which encompass from insurance (e.g., Ryals and Knox, 2005), credit card (e.g., Singh et al.,2013) until peer-to-peer lending (Machado & Karray, 2022b). The second highest group was in B2C settings (30%) outside the FSI, that have several industries, such as e-commerce (e.g., Ruch & Sackmann, 2009), telecommunication (Homburg et al., 2009), until airlines (Wangenheim & Lentz, 2005) . While the remaining of the studies (26%) was conducted on B2B settings, such as medical instrument provider for hospitals (Hai-wei et al., 2006) and commodities company (Yun & Yan, 2013).

In terms of application area, the studies conducted on risk-adjusted revenue primarily aimed at finding the optimal customer portfolio composition to maximize returns. Figure 6(B) shows that "customer portfolio management" is one of the dominant keywords in the area of research. This application area of the studies covered various industries such as the financial services industry (Dhar and Glazer, 2003; Buhl & Heinrich, 2008; Homburg et al., 2009; Albadvi & Norouzi, 2013), the business-to-consumer (B2C) setting (Wangenheim and Lentz, 2005; Sackmann et al., 2010; Norouzi & Albadvi, 2016; Viviani et al., 2021), and the business-to-business (B2B) setting (Tarasi et al., 2011; Juhl & Christensen, 2013; Yun & Yan, 2013). In addition to these industries, the studies also explored other areas of application, such as identifying the most profitable customers in the credit card industry (Singh et al., 2013; So et al., 2014), adjusting product prices in e-commerce (Ruch & Sackmann, 2012), and also find more accurate risk-adjusted CLV metric (Ryals, 2003; Singh & Singh, 2016; Petersen & Kumar, 2015; Machado & Karray, 2022a).

2.3.4 Methods to Incorporate the Customers' Risk in CLV Models

The incorporation of customers' risk into customer valuation is dependent on the specific application area of research. Many studies that apply financial portfolio concepts to customer portfolio optimization utilize mean-variance analysis to determine the optimal customer portfolio composition. To arrive at the best customer portfolio composition, these studies typically segment customers based on demographic features such as their profession, level of education, homeownership status, marital status, and employment status (Buhl & Heinrich, 2008; Homburg et al., 2009; Tarasi et al., 2011; Sackmann et al., 2010; Viviani et al., 2021; Machado & Karray, 2022b) before use it to calculate or predict the best customer portfolio composition. Customers' segmentation is necessary because the behavior of individual customers can be uncertain, while the behavior of a group consisting of a sufficient number of customers can be more predictable (Yun & Yan, 2013).

For instance, Tarasi et al. (2011) analyzed the variability in a customer portfolio, predicted the similarity of different market segments (i.e., automotive, consumer goods, food and beverage, etc), and explored the use of market segment weights in an optimized portfolio. They then used mean-variance analysis to create the best frontier of customer portfolio composition. Similarly, Viviani et al. (2021) segmented customers using the Hidden Markov Model (HMM) and used mean-variance analysis to construct an efficient customer portfolio while considering the customers' segment switching probability as the risk. In addition to mean-variance analysis, other methods are utilized to determine the optimal customer portfolio, such as particle swarm optimization (Yun & Yan, 2013).

The other studies with different application area, employ different approach to incorporate the customers' risk. For example, Ryals and Knox (2005), who employ a mathematical approach to directly calculate risk-adjusted CLV. So et al. (2014) uses statistical approach to build a scorecard predictive model and use them to identify more profitable customer. Albadvi and Norouzi (2013) cluster customers using the RFM model and then use a mathematical approach to calculate customers' risk. They estimate the risk-adjusted CLV associated with each segment using Pareto/NBD modeling. Singh et al. (2013) and Singh and Singh (2016) employed Data Envelopment Analysis (DEA) to compute a measure of RAR in a credit card company and CLV from a RARFM model, respectively.

2.3.5 Machine Learning in risk-adjusted CLV studies

From the studies found, only four studies use ML algorithms to predict the risk-adjusted CLV. The first study that apply ML algorithm in their work is presented by Wangenheim & Lentz (2005). They use multiple linear regression to predict the customer revenue after adjusted by the customers' risk and use the predicted value to build the customer segmentation using K-Means. Homburg et al. (2009) also partially utilized ML algorithms, a regression tree, to segment their customer data where the result is further used for customer portfolio optimization using Mean-Variance. In a more recent study, Machado and Karray (2022b) incorporated ML techniques into their model for predicting customers' RAR in the financial lending industry. To assess risk, they combined the PD of lenders obtained through logistic regression as one of their sources of customer's risk. They used the credit rating to predict the PD of the lender and employed two different methods to encode the customer credit rating, based on the Weight of Evidence (WOE) method and normalized interest rate, and compared the results. All of the previous approaches represent a novel attempt to integrate multiple sources of risk in CLV calculations, highlighting the potential benefits of combining traditional statistical methods with ML techniques.

And finally, the subsequent study by Machado & Karray (2022a) is the first study that fully utilized ML to predict the risk adjusted CLV. They propose a hybrid ML framework to predict the RAR value of customers' segment. The hybrid framework here refers to combine two different ML algorithm, usually supervised and unsupervised learning to predict a value. Hybrid models can offer better predictive results because they can handle high dimensional datasets more efficiently and combine individual models' best characteristics, hence mitigating their weaknesses. In their study, Machado and Karray compared the results of the individual models and hybrid models using customer data from the peer-to-peer lending industry. The hybrid model first clusters the customer data using unsupervised learning and then uses the results as a variable to predict the customer RAR value using supervised learning. The authors compared different combinations of clustering and predicting algorithms, such as k-Means++, k-Means random, DBSCAN, clustering with pre-determined variables for the clustering algorithms and Adaptive Boosting (AB), Gradient Boosting (GB), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) for the predicting algorithms. The experiments conducted in their research show that the hybrid ML algorithms outperform individual models in terms of predictive power and processing time for most frameworks. The combinations of DT, RF, or GB with k-Means++ exhibit higher efficiency than the individual models. Therefore, Machado and Karray's study demonstrates the potential of hybrid ML models in risk-adjusted CLV prediction and provides important insights for future studies in this area.

2.3.6 The customers' risk

The SLR study identified nine different sources of risk that customer might carry, with the most commonly used risks being the volatility of customers' income, beta risks, and customers' churn. The volatility of customers' income is a readily measurable risk that is available in all industry settings, which explains its prevalence in the literature. For example, Wangenheim and Lentz (2005) incorporated the volatility of customers' transactions as a non-systematic risk in their calculations to account for the limitations of the CAPM model in disregarding non-systematic risk. They also highlighted that when valuing an asset, the expected value is discounted by a factor influenced by the degree of risk associated with the asset, such as its past or future expected volatility. Therefore, the first metric to be tested in this context is the variance in customer transactions over time or the volatility of income from the customer, which could be defined as:

$$CVAR_i^j = \frac{1}{p-1}\sum_{t=1}^{p}(RV_{it} - CRV_i^j)^2 \qquad 2.1$$

Where p is is equal to the number of periods with $RV_{it} > 0$, c = 1,…, j. RV_{it} represent the revenue of customer i in period t, and CRV represent the conditional revenue of customer. The other study that use volatility of income as customers' risk is proposed by Homburg et al. (2009). They considered the volatility of customer income over a certain period and the likelihood of a customer changing their segment as risks to be factored in. Singh et al. (2013) proposed the inclusion of the volatility of multiple income sources as one of their multiple risks. They defined volatiltiy of income as:

$$\sigma = \sqrt{E(X^2) - (E(X))^2} \tag{2.2}$$

where E(X) is the expected average value of X. And finally, Machado & Karray (2022a) also use the volatility of income as the risk to predict the PD of a customer in FSI (Equation 2.2).

The second most commonly used risk factor identified in the SLR study is beta risk. As previously mentioned, most of the studies found were derived from financial portfolio theory (CAPM and MPT) and CAPM solely incorporates systematic risk, which is indicated by beta (Wangenheim & Lentz, 2005; Buhl & Heinrich, 2008). Beta is a metric that reflects the volatility of an asset's price fluctuations in comparison to a benchmark. Specifically, it denotes the ratio of the covariance between an asset and the market to the market's variance. If an asset's beta value is positive, the stock price is expected to increase and decrease in conjunction with the market; however, if the beta value is negative, the stock price is projected to rise (or fall) when the market declines (or rises) (Wangenheim & Lentz, 2005). A stock with a beta of 1.0 will have returns that track the market, whereas a stock with a beta of 1.5 is anticipated to move 1.5 times more than the market. Consequently, high beta shares are preferred in rising stock markets, while low beta shares are preferred in falling stock markets (Ryals L., 2003). This idea is then extended to the customer portfolio, where the beta value for each customer or customers' segment is computed and utilized as a measure of customers' risk.

To calculate the beta value of customers' segments, it is necessary to define the market portfolio, which consists of all available assets, with each asset held in proportion to its market value relative to the total market value of all assets. However, determining a single market portfolio for all companies is difficult or often not feasible (Buhl & Heinrich, 2008; Tarasi et al., 2011). Therefore, most studies define the market portfolio to calculate beta values as the company's current customer base (Dhar & Glazer, 2003; Wangenheim & Lentz, 2005; Buhl & Heinrich, 2008; Tarasi et al., 2011; Albadvi & Norouzi, 2013; Singh et al., 2013; Machado & Karray, 2022b). All of these studies use the same formula to calculate beta value:

$$\beta = \frac{cov(\varphi_{ct}, \varphi_{mt})}{var(\varphi_{mt})} \qquad 2.3$$

where φ_{ct} and φ_{mt} are the return for customer c and market m, respectively, at a given time period t. While the formula to calculate the discount rate might differ between studies, for example Dhar & Glazer (2003) and Wangenheim & Lentz (2005) define the discount rate as:

$$d = \psi_m \, x \, \beta \qquad 2.4$$

Where ψ_m represents the expected rate of return of the market. On the other hand, Buhl & Heinrich (2008) expand the formula to calculate the discount rate by adding the "risk free asset" or customer with the lowest risk in the formula, such as:

$$d = \psi_f + \beta \, (\psi_m - \psi_f) \qquad 2.5$$

Where ψ_f represent the minimum expected rate of return, which correspond with customer with the lowest risk and ψ_m represents the expected rate of return of the market.

The third most prevalent risk identified in the review was customer churn, which can be a risk factor in any industry and is now considered a major challenge for many companies (Singh & Singh, 2016). Customer churn refers to the probability that a customer will no longer be active or loyal to a company. Singh and Singh (2016) caution that relying solely on a customer's past purchases as a predictor of future behavior can be misleading, as it may incorrectly assume that the customer is still engaged with the company when in reality they may have already churned. This can lead companies to waste resources and money on customers who are no longer interested in their products or services. Hai-wei et al. (2006) note that traditional models like recency, frequency, monetary (RFM) often only use the recency of a customer's last purchase to estimate churn risk, but this may not accurately reflect the true risk of churn as it ignores variations in customer purchase cycles. To address this, they propose a more comprehensive definition of churn risk that takes into account the time elapsed since a customer's last purchase as well as their overall purchase history and therefore they define churn risk as :

$$Churn\ Risk = \frac{Recency}{\sqrt{\sigma_n}} \qquad 2.6$$

where σ_n represent the variance of interpurchase time. Ryals (2003) and Ryals and Knox (2005) use the probability of customer retention as their source of risk. They estimate the customer retention rate from the managerial knowledge on their customer relationship and use that risk to calculate the risk-adjusted CLV.

Several studies have identified different types of risks that are specific to particular industries. For example, in the financial services industry, such as credit card or loan providers, PD was used as a type of risk (Singh et al., 2013; So et al., 2014; Machado & Karray, 2022a). PD stands for the probability that a customer will default on their credit card or loan payments. Each financial institution has its own definition of when a customer is categorized as defaulting on payments, with most banks defining default as non-payment of minimum balance for a period of 90 days (Singh et al., 2013). Other source of risk were observed in other industries, for instance, in the insurance industry, customer claim risk and relationship risk were identified as types of risk (Ryals & Knox, 2005, 2007). For e-commerce, product returns and payment risk were identified as significant risks (Petersen & Kumar, 2015; Ruch & Sackmann, 2012).

2.3.7 The Time Horizon and Data Observed

The duration of observation or data used to calculate CLV varied significantly among the selected studies. Sackmann et al. (2010) had the shortest observation period, utilizing only 39 weeks of surveys to analyze customer behavior and segment customers. On the other hand, Buhl & Heinrich (2008) had the longest data observation period, utilizing 10 years of customer income data to predict the CLV. The duration of observation data was correlated with the industry in which the study was conducted. For example, Homburg et al. (2009) used different time ranges for each sector in their study. They define a period of observation as one quarter of a year for the telecommunications industry and one year for the banking industry. In general, the B2C sector tended to have a shorter time frame for data observation, while the FSI sector had the longest. The average time frame for B2C, B2B, and FSI was found to be 2.85, 3.5, and 4.9 years, respectively.

The number of customers observed in the studies also varied depending on the industry and method used. The smallest number of customers observed was 10 in Ryals & Knox's (2007) study on an insurance company, while Machado & Karray (2022a, 2022b) used the largest amount of data with a total of 2 million cases from 8 years of data. Ryals & Knox (2007) were able to use a small amount of customer data by employing managerial knowledge to determine the customer's risk value and then calculating the risk-adjusted CLV using mathematical approaches. Meanwhile, Machado & Karray (2022a, 2022b) used ML algorithms to predict customers' risk, which required more data to achieve better accuracy. The detailed information on the duration of the time period and number of customers used in each study can be found in Table 4.

2.4 Summary

This chapter provides an overview of ML and XAI in general, as well as their specific application in the context of customer valuation. As explained, ML is a potent tool for analyzing data and developing accurate predictive models. Meanwhile, XAI aims to provide interpretability and transparency to the decision-making process of ML models. While the application of ML is diverse, its application in predicting CLV has gained significant attention in recent studies, as it has been shown to outperform probability models However, while the application of ML in customer valuation is diverse and extensive, the use of XAI in this field remains limited. Nonetheless, XAI has been applied in other areas of the business field related to customer valuation, such as churn prediction, customer turnover, propensity to buy a product, and augmented cross-selling, among others.

This study also conducted a SLR to explore and evaluate the integration of customers' risk in customers' value calculation. The review identified 22 papers that integrated customers' risk in their customers' value calculation. The majority of these studies were based on financial portfolio theory, with a focus on finding the optimal customers' value composition to maximize the company's return. Another approach used was the MSR, which considers various risks to achieve a more accurate CLV metric and a better understanding of customers' value. The review also identified various industry settings and domains where these studies have been applied, which can be grouped into three main categories: FSI, B2C, and B2B.

The SLR study also identified the most significant types of customers' risk in calculating CLV across industries. The findings suggest that the most commonly used risk factor is the volatility of income from customers. This risk factor is easily measurable and applicable to all industry settings. Moreover, in non-contractual settings, income from customers tends to be more unpredictable, leading to a greater impact on CLV. The volatility of customer income is also closely associated with the second most commonly used risk factor, Beta value. Beta value measures volatility in asset price fluctuations relative to a benchmark. In the context of this study, Beta value reflects the volatility of customer income fluctuations in relation to the entire customer base. Beta value is widely used in studies seeking to identify the optimal customer portfolio. Customer churn emerged as the third most frequently used risk factor in calculating risk-adjusted CLV. This type of risk is applicable across all industry settings. However, it poses a particular challenge in non-contractual settings, where customers may silently churn, making it difficult to predict.

The SLR study also aimed to investigate the various methods used to incorporate customers' risk. The most prevalent approach in prior studies has been the mean-variance method. Typically, studies segment customers based on their financial and demographic characteristics and then utilize mean-variance optimization to identify the optimal customer portfolio. However, other methods have also been employed to assess the impact of customers' risk on customers' value. For instance, some studies have employed mathematical and DEA techniques, while others have utilized statistical approaches and ML algorithms to predict risk-adjusted customers' value. The application of ML-based methods in this field is gaining popularity, with several initial studies using ML algorithms in their process, such as customer segmentation and predicting customer probability of default. Moreover, there is a recent study that fully utilized an ML algorithm to predict the risk-adjusted revenue of customers.

This study also aims to examine the duration of observation and the amount of data used in previous studies, as these factors are highly associated with the industry setting and the method used in the study. The findings indicate that in previous studies, the B2C sector tended to have a shorter time frame for data observation, while the FSI sector had the longest. Specifically, the average time frame for B2C, B2B, and FSI was found to be 2.85, 3.5, and 4.9 years, respectively. Moreover, the amount of data used in previous studies varied significantly, ranging from only using the data of 10 customers to exploring the data from 2 million cases. These variations in the duration of observation and amount of data used should be taken into account when interpreting and comparing the results of different studies.

2.4.1 Research Gap

Despite the growing body of literature on customer valuation, there are still several research gaps that need to be addressed. Firstly, the studies in the field of customer valuation that integrate customers' risk have mostly been conducted in the FSI with limited representation from other industries. For instance, there is a lack of research in the B2C sector, which has different characteristics and challenges compared to FSI. Furthermore, there is only one study conducted in the telecommunication industry that explores the integration of customers' risk in customer valuation. This research gap is significant as the telecommunication industry is unique in its customer behavior, which is characterized by high customer acquisition cost and low switching costs. The application of customer valuation in the telecommunication industry, especially in predicting customer lifetime value, can provide valuable insights to optimize customer acquisition and retention strategies.

Secondly, while ML has been increasingly used in customer valuation studies, there is a lack of studies that fully utilize ML algorithms. According to Table 2, most studies focus on calculating the impact of customers' risk on CLV, while only a small portion employ prediction method to forecast the impact of customers' risk. Moreover, Table 4 shows that despite the growing popularity of ML, only two studies fully utilize this approach to predict risk-adjusted customers' value. This presents an opportunity for researchers to explore the potential of ML in predicting risk-adjusted customers' value. With the availability of big data and high computational power, ML has the potential to offer powerful predictions regarding customer behavior and risk. Thus, a comparative analysis between ML and traditional methods for calculating risk-adjusted customers' value would provide valuable insights into their effectiveness and potential applications across various industries. Further research in this area could pave the way for more accurate and precise CLV calculations, ultimately helping organizations to make informed decisions regarding customer acquisition and retention strategies.

Thirdly, there is a significant research gap in the application of XAI in customer valuation studies, with only one study found in this field. Moreover, there is no study found that has applied XAI to risk-adjusted revenue prediction. The gap in XAI research in customer valuation is crucial as the explainability of ML models in customer valuation is increasingly important due to the interpretability, regulatory, and ethical considerations. Furthermore, the lack of XAI applications in risk-adjusted revenue prediction is an important gap as it is an essential metric for customer valuation that integrates customers' risk. The integration of XAI in risk-adjusted revenue prediction can enhance the interpretability of ML models and provide more accurate and reliable customer valuation assessments.

In summary, the current literature on customer valuation presents several research gaps, including the lack of representation from other industries, the limited use of ML, and the lack of studies utilizing XAI. This thesis addresses these research gaps, helping to improve the accuracy and reliability of customer valuation methods and provide more actionable insights for decision-making in various industries.

3 Methodology

This chapter presents the methodology that was employed to achieve the research objective and address the research question and sub-research questions, as established in Section 1.3. The initial segment outlines the comprehensive methodological approach that was adopted in this study. Subsequently, the chapter expounds on the fundamental aspects of the methodology that are pertinent to the topic investigated in this thesis, such as customer valuation, ML algorithms, and XAI.

3.1 Research Design Methodology

To achieve the research objective of developing and evaluating an artifact to provide a quantifiable measure of the risk-adjusted expected future value of a customer in the telecommunication industry, the DSRM as proposed by Peffers et al. (2007) was selected as the primary methodology. This selection was made based on its alignment with the research objective. The artifact that was developed is a ML model, which aimed to predict the risk-adjusted CLV of customers in the telecommunication industry. DSRM provided a structured framework that ensured a systematic approach to artifact development, testing, and evaluation. It emphasized the importance of rigor, relevance, and utility in creating artifacts that addressed real-world problems. Using DSRM methodology ensured that the development of the ML model was grounded in a well-defined and systematic research process.

In addition to DSRM, the Cross-Industry Standard Process for Data Mining (CRISP-DM) was incorporated as a sub-methodology to guide the development and evaluation of the ML model. The CRISP-DM methodology is a widely accepted, industry-independent process model for data mining and ML model development. Despite its introduction over two decades ago, CRISP-DM remains the standard in aligning data mining with business goals (Schröer et al., 2021). By integrating CRISP-DM, the process of developing and evaluating the ML model could be effectively guided, ensuring that all necessary steps were considered and executed systematically. Combining DSRM and CRISP-DM allowed for a comprehensive research methodology that covered both the design and evaluation aspects of the artifact, as well as the data-driven and technical components of the ML model. This integrated approach ensured that the developed artifact aligned with the research objective, was grounded in theoretical foundations, and followed industry best practices for data mining and predictive modeling.

As mentioned in Section 1.5, DSRM consisted of six steps, namely problem identification and motivation, defining objectives for a solution, design and development, demonstration, evaluation, and communication (Figure 1). Furthermore, CRISP-DM is comprised of six distinct phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Wirth & Hipp, 2000). These phases are iterative in nature and allow for movement between phases as necessary. Figure 7 illustrates the flow of each of the CRISP-DM phases. To operationalize the research methodology, the integration of the phases of DSRM and CRISP-DM was used in a structured and iterative approach. The six steps of DSRM guided the overall research design and artifact development, while the six phases of CRISP-DM were incorporated as a sub-methodology to guide the development and evaluation of the ML model.

Each step of CRISP-DM was integrated with the relevant DSRM steps to ensure alignment and completeness. Specifically, the business understanding phase of CRISP-DM was integrated with the problem identification and motivation and defining objectives for a solution phase of DSRM, respectively. The data understanding, data preparation, and modelling phases of CRISP-DM were integrated with the design and development phase of DSRM. The evaluation phase of CRISP-DM was integrated with the demonstration and evaluation phases of DSRM. During the evaluation phase of CRISP-DM, the results were evaluated against the defined objective. Similarly, in the demonstration phase of DSRM, the utility, efficacy, and usability of the artifact in a real-world context were demonstrated, including its effectiveness and efficiency in addressing the research problem. In the evaluation phase of DSRM, the effectiveness and efficiency of the artifact were evaluated, and any limitations or improvements were identified. This phase aligned with the evaluation phase of CRISP-DM. The deployment phase of CRISP-DM, which deploys the ML model in a production environment, monitors its performance over time, and assesses its impact on the business, was not used in this thesis. This study only aimed to develop the model and assess its performance without deploying it into the system.

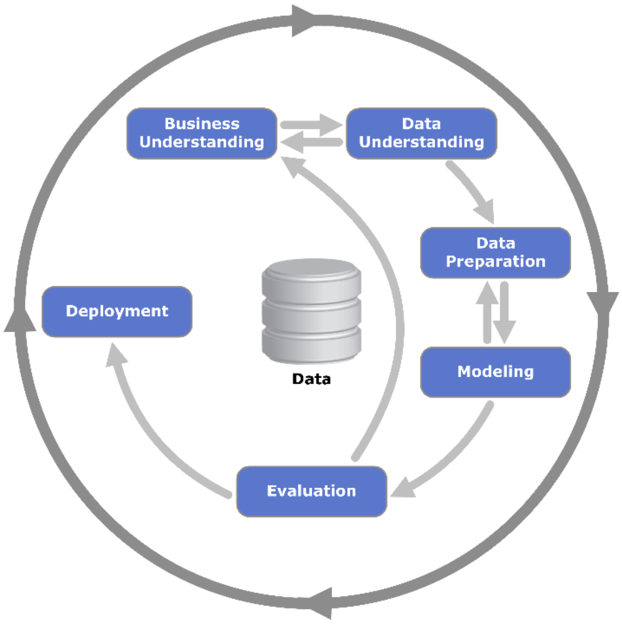


Figure 7 CRISP-DM Methodology (Jensen, 2012)

3.1.1 Problem Identification and Motivation

The problem identification has already been stated in the Section 1.2 of this thesis. As previously stated, accurately estimating CLV is crucial for organizations to make informed decisions regarding customer acquisition and retention strategies. CLV provides valuable insights into the profitability of current and potential customers, allowing companies to prioritize their marketing efforts and allocate resources effectively. However, incorporating customers' risk into CLV calculation remains a challenge, with limited studies conducted using traditional methods. On the other hand, ML has emerged as a tool for analyzing and building prediction models using big data. The application of ML for CLV prediction has gained attention in recent studies as it has shown that ML-based CLV models can outperform traditional models. Therefore, there is a need to explore the potential of utilizing ML to develop a new metric to predict the risk-adjusted CLV of customers.

3.1.2 Define objective of a Solution

The objectives of this study are outlined in Sections 1.3 and 1.4 of this thesis. The primary objective is to develop a novel metric for predicting risk-adjusted CLV using ML techniques. Existing research indicates that the majority of studies that incorporate risk in CLV calculation are conducted in the FSI using traditional methods. Thus, this study aims to address this gap by exploring the potential of ML to develop a new metric for predicting risk-adjusted CLV in the telecommunications industry. By conducting this research in a less studied field, such as the telecommunications industry, this study contributes to the existing knowledge on risk-adjusted CLV and enhances understanding of the application of ML in this domain. Additionally, the literature review also highlights a lack of studies that utilize XAI to predict risk-adjusted CLV. Hence, the second objective of this study is to develop an XAI model for predicting risk-adjusted CLV. This further enhances the interpretability of the proposed metric, contributing to the wider adoption of ML-based CLV models.

3.1.3 Design and Development

In the third step of the DSRM process, an artifact is created as a solution to the problem, which may take the form of models, concepts, or other methods. According to Peffers et al.(2007), an artifact can be broadly defined as "any designed object in which a research contribution is embedded in the design." In this thesis, the artifact refers to the design and development of a ML model to predict the risk-adjusted CLV. The design and development steps of DSRM will be following the data understanding, data preparation and modelling phase of CRISP-DM method.

The data understanding phase is crucial in the CRISP-DM methodology as it lays the foundation for the subsequent phases of data preparation and modeling. During this phase, the data is collected and explored to gain a better understanding of its characteristics, quality, and structure. This involves identifying potential data sources, examining the data's distribution and summary statistics, and detecting potential outliers or missing values. Additionally, data understanding may also entail exploring the relationships and dependencies between different variables and features within the dataset. This phase is covered in Chapter 4.

Further, the raw data will be transformed into a format suitable for modeling in data preparation phase. This process often includes data cleaning, selection, and transformation. Firstly, the data will be cleaned to handle errors, missing data, and inconsistencies. Then, a relevant subset of the data will be selected for inclusion in the model development. Finally, the data will be transformed into a format suitable for the chosen modeling technique. This phase is crucial for ensuring the quality of the data and the success of the subsequent modeling phase. The outcome of this phase is a prepared dataset that is ready for modeling, testing, and evaluation. This phase will be explained in Chapter 4.

Finally, in the modelling phase, the model is designed and developed based on the available data. The first step is to select an appropriate modelling technique that suits the business problem and the data available. This selection process can be guided by the models used in the literature review. The performance of the model is then evaluated using a separate subset of data, known as the validation, set. This process is repeated until an optimal model is achieved. Feature selection is another crucial aspect of the modelling phase, which involves identifying the relevant features or variables that have the most significant impact on the model's predictive power. The objective is to include only the most relevant features in the model to avoid overfitting or underfitting, which can reduce the model's performance. Additionally, the model's performance is tested on new and unseen data to assess its ability to generalize to new situations. This process is called model validation or testing, and it is essential to ensure that the model performs well on data that it has not seen before. The modelling and data preparation phases are iterative, and if the model's performance is not satisfactory, the process can be restarted until an optimal model is obtained. Figure 7 illustrates this iterative process, where the modelling and data preparation phases are not unidirectional.

3.1.4 Demonstration and Evaluation

According to Peffers et al. (2007), the demonstration step of DSRM include demonstrate the practical application of the developed solution. This can take the form of experiments, simulations, or case studies. And the evaluation involves in testing the artifact created in the previous step and evaluating its effectiveness in solving the identified problem. This step focuses on validating the effectiveness, efficiency, and suitability of the artifact in meeting the desired goals and objectives. The artifact is tested using real-world data to determine its performance and functionality. The results are then evaluated and compared with the desired outcomes and expectations to determine the extent to which the artifact meets the specified criteria. These two steps are aligned with the evaluation phase of CRISP-DM methodology.

3.1.5 Communication

The final stage of the research aims to disseminate the findings and outcomes of the study. This is achieved through a comprehensive report that presents the research process, challenges, and artifacts, as well as novel insights and relevant information that can help stakeholders understand the research problem and its solutions. The report targets organizations, researchers, and audiences who are interested in the topic and provides recommendations for future research and practical applications.

3.2 Analytical Methods

This section outlines the specific analytical methods that will be employed in this thesis, including the method to calculate CLV, the specific ML methods that used, and the XAI methods that applied. The section provides a detailed overview of each of these methods, including their theoretical basis, strengths, and limitations, as well as their specific applications to the research study. By outlining the analytical methods that used in this study, this section aims to provide a clear and comprehensive understanding of the analytical framework that guides the research, and to demonstrate the rigor and validity of the study's findings.

3.2.1 Customer Lifetime Value

There are various methods and formulas that can be used to calculate CLV, and the appropriate approach will depend on the nature of the business and the available data. However, the two fundamental steps for assessing CLV are: (a) forecasting the net cash inflows the company from the customer over a period of time, and (b) determining the current value of that series of cash flows. Hence, according to Berger and Nasr (1998) the general CLV formula is (Berger & Nasr, 1998) :

$$CLV = \sum_{i=0}^{n} \pi(t) \times \frac{1}{(1+d)^t}, i = 1, \ldots, n \tag{3.1}$$

Where $\pi(t)$ is represents the cash flow generated from customer i in period t, and d is the discount rate.

Another method to calculate CLV is proposed by Gupta & Lehmann (2003), where they suggested an alternative approach to compute CLV that draws on the discounted cash flow method in finance. Nevertheless, there are two significant distinctions between the two methods. Firstly, CLV is generally defined and estimated at an individual customer or segment level. This enables a distinction to be made between more profitable customers as opposed to just looking at overall profitability averages. Secondly, unlike finance, CLV incorporates the potential for a customer to defect to a competitor in the future, making it more comprehensive in nature.

$$CLV = \sum_{t=0}^{T} \frac{(p_t - c_t)r_t}{(1+d)^t} - AC \tag{3.2}$$

where p_t it the price paid by a consumer at time t, $c_t$is the direct cost of servicing the customer at time t, d is discount rate or cost of capital for the firm, r_t is probability of customer repeat buying or being "alive" at time t, AC = acquisition cost, and T = time horizon for estimating CLV. The formula demonstrates that the calculation of CLV is not solely based on the revenue generated from a customer, but also heavily influenced by the probability of repeat purchases, which is commonly referred to as the retention rate. This highlights the significance of customer retention in determining the customer's value and the risk associated with their retention, which could ultimately impact the CLV. Further, Gupta and Lehmann (2003, 2005) also demonstrated that under the assumption of constant margins $(p_t - c_t)$ and retention rates over time, and with an infinite time horizon, the expression for CLV can be simplified to the following equation:

$$CLV = \sum_{t=0}^{\infty} m \frac{r}{(1+d-r)^t} \qquad 3.3$$

Hwang et al. (2004) proposes a new model for measuring lifetime value that takes into account both customer defection and cross-selling opportunities in a business. Furthermore, unlike the existing model that focused on financial contribution estimated from past history of profit generation and converted the contribution to present value, the authors argue that their proposed model focuses not only on past profit contribution, but also on future financial contribution, potential profit generation of a customer, and expected service periods. Their prosed CLV model is formulated as follows:

$$CLV = \sum_{t_i=0}^{N_i} \pi_p(t_i)(1+d)^{N_i-t_i} + \sum_{t_i=N_i+1}^{N_i+E(i)+1} \frac{\pi_f(t_i)+B(t_i)}{(1+d)^{t_i-N_i}} \qquad 3.4$$

Where N_i represent the total service period of customer i and t_i represent service period index of customer i. The first part of the equation, the sum of $\pi_p(t_i)(1+d)^{N_i-t_i}$, represents the net present value (NPV) of the past profit contribution of the customer, calculated by summing up the profit contribution of customer i at period (t_i), multiplied by the interest rate factor $((1+d)^{N_i-t_i})$, which transforms the past profit into the present value. The future cash flow is computed by adding the expected future profit and the potential benefits (B) that the customer could bring during the expected service period i (t_i). The potential benefits or value refer to the expected profits that a company can obtain from a customer by providing additional services or cross-selling.

Several studies extend the general CLV formulas by incorporate customers' risks in the discount rate (d) calculation. Ryals & Knox (2005) used two criteria to calculate the discount rate, the risk of customers and market $(\frac{R_c}{R_m})$ and Weighted Average Cost of Capital (WACC[5]). They define the customers' risks as the weighted customer credit rating for individual customers and its average for the entire portfolio representing the market risk (Machado & Karray, 2022a).

$$d = \frac{WACC \times R_c}{R_m} \tag{3.5}$$

(Dhar & Glazer, 2003) applied the CAPM theory and include the systematic risk, Beta value (β) in the discount rate calculation, where the formula became:

$$\beta = \frac{cov(\varphi_{ct}, \varphi_{mt})}{var(\varphi_{mt})} \tag{3.6}$$

$$d = \psi_m \, x \, \beta \tag{3.7}$$

where φ_{ct} *and* φ_{mt} are the return for customer c and market m, respectively, at a given time period t and ψ_m represents the expected rate of return of the market. Finally, Buhl & Heinrich (2008) expand the formula to calculate the discount rate by adding the "risk free asset" or customer with the lowest risk in the formula, such as:

$$d = \psi_f + \beta\,(\psi_m - \psi_f) \tag{3.8}$$

Where ψ_f represent the minimum expected rate of return, which correspond with customer with the lowest risk and ψ_m represents the expected rate of return of the market. The detailed definition of beta value could be find in the Subsection 2.3.6.

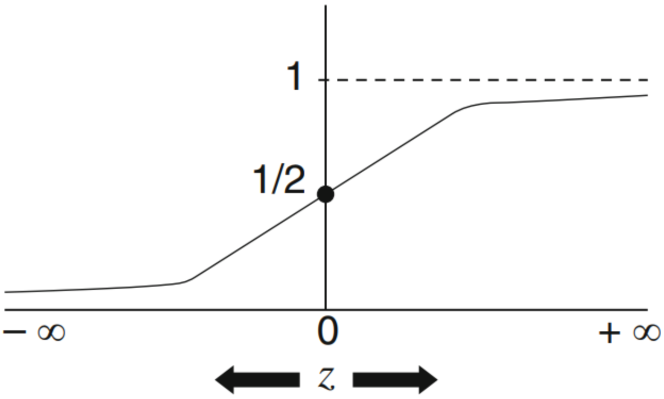
3.2.2 Machine Learning Algorithms

3.2.2.1 Logistic Regression

Logistic regression (LR), also referred to as the logit model or logistic model, is a statistical method used to assess the association between a categorical dependent variable and several independent variables. This approach estimates the likelihood of an event occurring by fitting data to a logistic curve. (Kleinbaum et al., 2002). Figure 8 displays the logistic function

[5] A financial metric used to calculate the average cost of financing a company's operations by taking into account the proportion of debt and equity and their respective costs.

that underlies the mathematical framework of the logistic model. The function, denoted as $f(z)$, illustrates the probability of an event occurring by fitting data to a logistic curve. It is represented by the equation $\frac{1}{1+e^{-z}}$, where z represents the input values. The plotted graph illustrates the range of values that $f(z)$ can take, as z varies from 1 to +1.



$$f(z) = \frac{1}{1+e^{-z}} \qquad 3.9$$

Figure 8 Logistic function (Kleinbaum et al., 2002)

Examining the left side of the graph, it is notable that the logistic function $f(z)$ takes the value of 0 when z tends towards $-\infty$. In contrast, on the right side, as z approaches ∞, $f(z)$ approaches 1. Regardless of the specific value of z, $f(z)$ consistently ranges between 0 and 1. This inherent characteristic of the logistic function, tailored to accommodate probabilities, ensures that any risk estimate derived from it lies within the interval of 0 to 1. The logistic model is purposefully designed to guarantee that any estimated risk will constantly reside within the bounds of 0 and 1. Consequently, the logistic model never yields a risk estimate surpassing 1 or falling below 0. This sets it apart from alternative models, thereby positioning the logistic model as the preferred choice when seeking to estimate probabilities (Kleinbaum et al., 2002).

The allure of the logistic model further stems from the shape of its corresponding logistic function. As depicted in the Figure 8, when moving from z approaching negative infinity towards the right, $f(z)$ initially maintains proximity to zero, subsequently experiences a rapid ascent, and eventually approaches but never reaches 1 as z increases. This pattern generates an elongated and distinctively S-shaped curve, contributing to the unique representation of the logistic model. The sigmoid or S-shaped nature of the logistic function holds particular appeal, especially when the variable z is interpreted as an index that consolidates the contributions of variables, while $f(z)$ represents the corresponding probability for a given value of z. In this context, the S-shape of $f(z)$ signifies that the influence of z on the variable remains relatively minimal for lower values of z until a certain threshold is reached. Subsequently, the probability escalates rapidly within a specific range of intermediate z values, but never reaches 1, indicating that the model is designed to estimate probabilities within the range of 0 to 1.

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \qquad 3.10$$

To establish the logistic model, the variable z is represented as the sum of a linear combination of independent variables X_1, X_2, up to X_k, where α and β_i are constant terms signifying unknown parameters. Conceptually, z acts as an index that amalgamates the values of the X_s. Substituting this linear sum expression into the logistic function, the resulting equation becomes:

$$f(z) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}} \qquad 3.11$$

To fully grasp the mathematical model, it will be explained within an epidemiological context. The logistic model operates within the framework of a generalized epidemiological study, where a group of subjects is observed with respect to independent variables X_1, X_2, up to X_k. Additionally, disease status is determined, denoted as 1 for "with disease" or 0 for "without disease." The objective is to utilize this information to depict the probability of disease development during a defined study period (e.g., from T_0 to T_1) for disease-free individuals, characterized by specific values of the independent variables X_1, X_2, up to X_k measured at T_0.

The probability of developing the disease, given the independent variables X_s, is represented as $P(D = 1|X_1, X_2, \dots, X_k)$. In the logistic model, this probability is defined as:

$$P(D = 1|X_1, X_2, \dots, X_k) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}} \qquad 3.12$$

The parameters α and the β_i in the model correspond to unknown values that necessitate estimation based on collected data concerning the X_s and the disease outcome (D) for the subject group. Consequently, if the values of the parameters α and β_i are known, along with the specific X_1 through X_k values for a particular disease-free individual, the logistic model can be employed to determine the probability of that individual developing the disease over a defined follow-up time interval.

For ease of notation, the probability statement $P(D = 1|X_1, X_2, \ldots, X_k)$ is succinctly denoted as $P(\boldsymbol{X})$, wherein the bold $\boldsymbol{X}$ serves as a compact representation encompassing the collection of variables X_1 through X_k. Therefore, the logistic model can be expressed as:

$$P(X) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}} \quad 3.13$$

3.2.2.2 Random Forest

Random Forest is a highly regarded ensemble learning method widely employed in both classification and regression problems. Introduced by Leo Breiman (Cutler et al., 2012), RF utilize the collective wisdom of multiple decision trees to produce a consolidated result. This algorithm extends the bootstrapping method proposed by Breiman in 1996 and forms a crucial component of ensemble algorithms. During the training phase, the RF algorithm generates numerous decision trees. Each tree randomly selects a subset of data through bootstrapping, a process where data is sampled with replacement. Furthermore, randomness is introduced at each decision split by selecting different subsets of features, resulting in a diverse set of models (Breiman, 2001; Cutler et al., 2012). Once multiple data samples are generated, these models are independently trained, and their predictions are aggregated, usually by taking their average or majority, to yield a more accurate estimate (Breiman, 2001). Figure 9 below shows the workflows of RF algorithm.

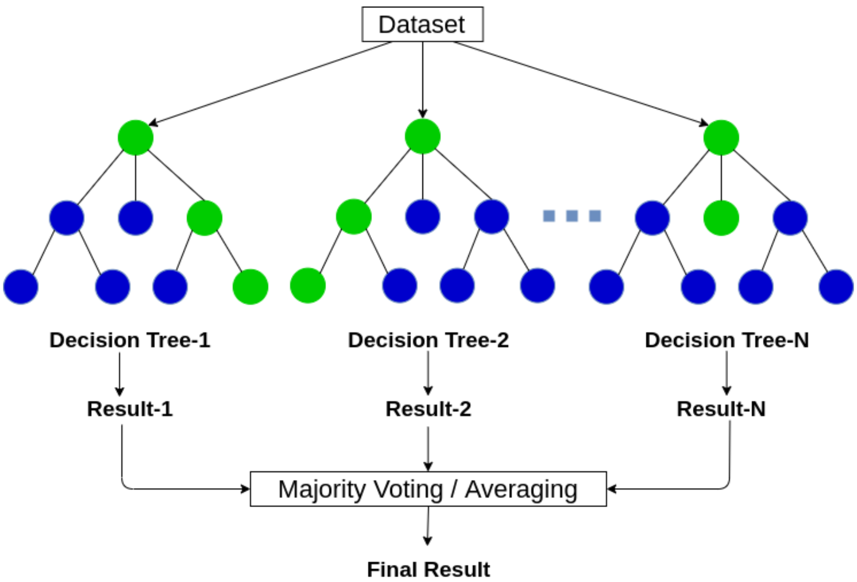


Figure 9 Random Forest (Sharma, 2020)

By incorporating randomness at both the data and feature levels, the algorithm generates an ensemble of diverse trees. With a large number of trees, the RF model benefits from the Strong Law of Large Numbers, which helps overcome the overfitting problem often encountered with single decision trees. Overfitting occurs when a model performs well on the training data but poorly on unseen data. The diversity of the tree ensemble, coupled with the aggregation process, allows the Random Forest model to achieve robustness and improved accuracy (Breiman, 2001).

A notable advantage of RF is their ability to handle both regression and multiclass classification problems effortlessly. They exhibit relatively fast training and prediction times and depend on only one or two tuning parameters, simplifying the model selection process. Moreover, RF provide measures of variable importance, which aid in feature selection and provide insights into the underlying relationships within the data. Additionally, RF are adept at handling missing values and outliers, further enhancing their versatility (Cutler et al., 2012).

However, there are certain drawbacks to be considered in RF algorithm. If the number of trees in the forest is not properly tuned, RF can be computationally expensive without significant performance gain. Therefore, selecting an appropriate number of trees is crucial to strike a balance between computational cost and model effectiveness (Oshiro et al., 2012). While RF provide some degree of interpretability through variable importance measures, they may not offer the same level of interpretability as simpler models like linear regression or single decision trees. It is essential to consider the trade-off between interpretability and predictive performance when choosing the appropriate modeling approach (IBM, n.d.).

In the case of the RF Regressor, a specific application of the RF algorithm, it is employed when the target variable is continuous. Operating under the same fundamental principles as the RF classifier, the RF Regressor differs in how predictions are aggregated. The final prediction is computed as the average of the predictions from all trees in the forest, reducing variance and mitigating the risk of overfitting, especially when the loss function is Mean Squared Error (MSE) (Cutler et al., 2012).

In conclusion, RF are a powerful ensemble learning method that effectively addresses classification and regression problems. By harnessing the collective strength of multiple decision trees, RF offer improved accuracy, robustness, and versatility. While they have advantages such as handling various data types, providing measures of variable importance, and mitigating overfitting, it is important to carefully tune the number of trees and consider the trade-off between interpretability and predictive performance. The RF Regressor variant specifically caters to continuous target variables, employing an averaging approach to enhance model performance.

3.2.2.3 XGBoost

Extreme Gradient Boosting (XGBoost), another ensemble ML algorithm, was first introduced by Chen and Guestrin in 2016. Similar to RF, XGBoost uses decision trees as its base learners, combining the predictions of several models to enhance robustness and improve generalization over a single model. However, unlike RF, which employs bootstrapping or bagging to enhance the performance of the base learners, XGBoost utilizes boosting algorithms as its learning strategy. Boosting is a sequential technique that combines a set of weak learners (in this case, single decision trees) to deliver improved prediction accuracy. It operates in a stage-wise manner, training weak learners in sequence, with each one aiming to correct the mistakes of its predecessors. Each model in the sequence is assigned a weight reflecting its contribution to the final prediction, which is a weighted sum of the predictions of the individual models (Hastie et al., 2009).

Following its namesake, "Extreme Gradient Boosting", XGBoost embraces the concept of a boosting method known as Gradient Boosting (GB). This technique, introduced by Friedman in 2001, uses a gradient descent algorithm to minimize errors in sequential models. The GB algorithm creates new models that predict the residual errors of prior models and combines these new error-predicting models to construct the final prediction. XGBoost refines the basic GB algorithm by introducing a regularization parameter to control overfitting, rendering it robust to noisy data and outliers (Chen & Guestrin, 2016). This unique feature allows XGBoost to be effectively employed as both a classifier and a regression algorithm.

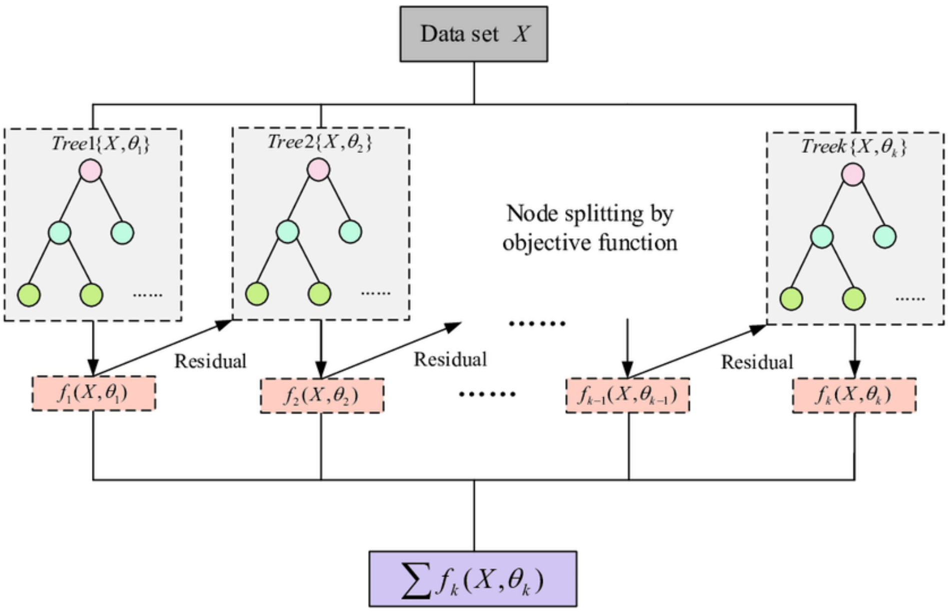


Figure 10 XGBoost Workflow (Guo et al., 2020)

Aside from its capability to prevent overfitting, XGBoost demonstrates high predictive power and outpaces other gradient boosting techniques in terms of speed. This advantage is due to its utilization of column block data structures in which the data is stored pre-sorted, enabling parallel learning and significantly reducing computation time. Moreover, it encompasses an out-of-core computation feature that optimizes memory usage, making it particularly well-suited for large datasets (Bentéjac et al., 2021). XGBoost can handle both linear and tree learners, learn feature interactions automatically, and includes built-in cross-validation at each iteration of the boosting process.

On the other hand, XGBoost also has several drawbacks. Firstly, to attain optimum performance, XGBoost requires meticulous tuning of parameters, which can be time-consuming. Secondly, XGBoost can still overfit the training data if the regularization parameter is improperly configured. Lastly, while it is more interpretable than some complex models, such as neural networks, it does not compare to the straightforwardness of simpler models like linear regression or decision trees.

3.2.2.4 CatBoost

CatBoost, derived from "Category" and "Boosting," is an implementation of gradient boosting algorithms that specifically designed to handle categorical variables and reduce the prediction shift that occurs during training phase of gradient boosting algorithm (Prokhorenkova et al., 2018). In CatBoost, categorical features are replaced by a numerical feature that signifies the expected target value for each category. Ideally, this numerical feature should be computed using a different dataset to avoid overfitting to the training data, but this is not always feasible. The procedure proposed in CatBoost calculates this new feature in a manner similar to the model-building process. For a given random permutation of the instances, the information from instances < i is used to compute the feature value of instance i. Subsequently, several permutations are carried out, and the obtained feature value for each instance is averaged, which aids in avoiding overfitting to the training data (Prokhorenkova et al., 2018; Bentéjac et al., 2021).

The primary methodology of CatBoost diverges from typical gradient boosting algorithms via its implementation of "ordered boosting." This method, driven by permutations, resolves the prediction shift commonly encountered in gradient boosting. This shift originates from the discrepancy between $F(x_i)|\ x_i$, representing a training instance, and $F(x_i)|\ x_i$ for a test instance x. It transpires because gradient boosting employs the same instances for estimating both the gradients and the models that minimize these gradients (Bentéjac et al., 2021). CatBoost proposes a solution wherein the gradients are estimated using a series of base models that exclude the instance in question from their training sets. This is initiated by introducing a random permutation in the training instances. The concept behind CatBoost is to build N base models for each of the M boosting iterations. The i-th model of the m-th iteration is trained on the first i instances of the permutation and is employed to estimate the gradient of the $i + 1$ instance for the $(m + 1)$th boosting iteration. To maintain independence from the initial random permutation, this process is repeated using s different random permutations. The implementation of CatBoost optimizes this process such that a single model is built per iteration, accommodating all permutations and models. The base models employed are symmetric trees or decision tables, which are grown by extending all leaf nodes level-wise using the same split condition (Bentéjac et al., 2021).

One of the key strengths of CatBoost is its robustness against overfitting. This feature is accomplished by employing oblivious trees, which use the same split for a feature across all instances. In conjunction with this, a regularization term is incorporated in the loss function, further preventing overfitting. Benchmark studies have also demonstrated the superior performance of CatBoost over other gradient boosting algorithms and random forests, especially with categorical data (Jhaveri et al., 2019). However, despite these strengths, CatBoost has its limitations. It may run slower than its counterparts for large-scale numerical datasets, and careful parameter tuning may be necessary for optimal performance.

3.2.3 Evaluation Metrics in Machine Learning Models

Evaluation metrics are essential in assessing the performance of ML models. They provide quantitative insight into the level of accuracy or error inherent in a prediction model, helping data scientists tune their models for optimal outcomes.

In classification tasks, some of the most commonly used evaluation metrics are:

- Accuracy: This is the simplest evaluation metric for classification, representing the ratio of correct predictions to total predictions. Accuracy is commonly used when target classes are well balanced.
- Precision, Recall, and F1 Score: These metrics are derived from the confusion matrix, a table layout that visualizes the performance of an algorithm (Powers, 2011). Precision measures the proportion of true positive predictions among all positive predictions. Recall, also known as sensitivity or true positive rate, measures the proportion of true positives among all actual positives. The F1

Score is the harmonic mean of Precision and Recall, providing a balance between the two metrics (Sokolova and Lapalme, 2009).

- Area Under the Receiver Operating Characteristic curve (AUC-ROC): The ROC curve plots the true positive rate against the false positive rate for various threshold settings, and the AUC quantifies the overall performance of the classifier. It provides a single scalar value representing expected performance (Fawcett, 2006).

In regression tasks, the following metrics are often utilized:

- Mean Absolute Error (MAE): It calculates the average of the absolute differences between the predicted and actual values. MAE is a linear score, meaning all individual differences are weighted equally (Willmott & Matsuura, 2005).
- Mean Squared Error (MSE): Similar to MAE, it calculates the average of the squared differences between the predicted and actual values. MSE gives more weight to larger differences (Willmott and Matsuura, 2005).
- Root Mean Squared Error (RMSE): The RMSE is the square root of the MSE. The RMSE is preferred over the MSE when large errors are particularly undesirable, as the RMSE is more sensitive to such errors (Willmott and Matsuura, 2005).
- Mean Absolute Percentage Error (MAPE): MAPE measures the average absolute percent error for each time period or data point. It is often used as a loss function for regression problems and in model evaluation because of its easy interpretability as percentage error (Myttenaere et al., 2016)
- R^2: Also known as the coefficient of determination, R^2 quantifies the proportion of variance in the dependent variable that can be predicted from the independent variable(s). It measures the strength of the relationship between the model and the dependent variable on a convenient 0 – 100% scale. (Chicco et al., 2021)

Each of these metrics presents a different perspective on the errors inherent in a prediction model and can be useful for diagnosing specific issues.

3.2.4 Explainable AI (XAI)

Just like mentioned in Section 2.1, the field of XAI can be categorized into two main strands of work: transparency design and post-hoc explanation (Lipton, 2018; Linardatos et al., 2021). As the interest to decipher the complexity of black-box models, the field of XAI has proposed a variety of post-hoc interpretability methods. These tools can be categorized into two broad types based on the level of their interpretability: global and local models. Global models, such as SHAP, aim to illustrate the overall behavior of a model by interpreting the average impact of features across all instances (Lundberg & Lee, 2017). Local models, exemplified by Local Interpretable Model-agnostic Explanations (LIME), focus on explaining individual predictions, providing a detailed understanding of the model's decision-making process for specific instances (Ribeiro et al., 2016).

3.2.4.1 Local Interpretable Model-agnostic Explanations (LIME)

LIME is a technique within the field of XAI that provides human-interpretable explanations of individual predictions made by complex ML models. It does this by approximating the local decision boundary around the point of prediction using a simpler model (e.g., a linear model), which can then be inspected for understanding (Ribeiro, Singh, & Guestrin, 2016). LIME works on the assumption that every complex model is linear on a local scale, despite being potentially nonlinear on a global scale. This makes it a model-agnostic method, meaning it can be applied to any ML model (Ribeiro et al., 2016). The simplicity of LIME's surrogate model ensures the interpretability of its explanations, allowing humans to understand and potentially trust the decisions made by the complex model.

The process employed by LIME involves the following steps: First, it select the set of features to explain, which can be all features in the dataset or a subset of particular interest. A grid of points is then established around the instance being examined, with its dimensions defined in a way that captures the local feature variation yet remains sufficiently representative of the instance's vicinity. Perturbed instances are then generated by introducing random adjustments to the feature values for each point in the grid. Following this, an interpretable model, such as a linear model or decision tree, is fitted to these perturbed instances. Finally, this fitted model is used to elucidate the prediction made by the black box model for the original instance (Ribeiro et al., 2016).

One of the strengths of LIME is that it allows users to understand and trust the predictions of complex ML models by providing local explanations. It also facilitates the detection and correction of biases, errors, and undesired behaviors in the models. However, one limitation of LIME is that it only provides local, rather than global, explanations (Ribeiro et al., 2016). Overall, LIME has proven to be a useful tool for demystifying the complex predictions made by black-box models and making ML more transparent, trustworthy, and accessible.

3.2.4.2 SHapley Additive exPlanations (SHAP) value

While LIME provides valuable local explanations for specific instances, there is a growing need for interpretability methods that can provide both global and local explanations. SHAP is a method that offers this dual perspective. SHAP, developed by Lundberg and Lee (2017), is built upon the foundation of Shapley values, a concept first introduced within the realm of cooperative game theory. In their original context, Shapley values quantify the contributions of individual players to a cooperative game. This concept of assigning contributions to individual players has been adapted and applied in the ML domain to determine the importance of individual features in predictive models. With SHAP, each feature's impact on a model's output is measured, enabling explanations for individual predictions to be formulated. Hence, SHAP allows for a more comprehensive understanding of model predictions, attributing each feature's influence on both a specific prediction (local interpretation) and across all predictions (global interpretation).

The primary objective of the SHAP method is to provide an explanation for the prediction made by a given instance, denoted as x, by quantifying the contribution of each feature towards that prediction. This explanation technique relies on the computation of Shapley values, which are derived from coalitional game theory. In the context of SHAP, the feature values associated with a particular data instance serve as players within a coalition. The Shapley values enable a fair allocation of the "payout" (i.e., the prediction) among the features involved. In the case of tabular data, a player can refer to an individual feature value. However, a player can also be a group of feature values. For example, to explain an image, pixels can be grouped to super pixels and the prediction distributed among them. A notable aspect introduced by the SHAP method is its representation of Shapley value explanations as an additive feature attribution method, akin to a linear model. This perspective establishes a connection between LIME and Shapley values. SHAP presents the explanation in the following manner:

$$g(z') = \emptyset_0 + \sum_{j=1}^{M} \emptyset_j z'_j \qquad 3.14$$

Where g is the explanation model, $z' \in \{0,1\}^M$is simplified input, M represents the number of simplified input feature and $\emptyset_j \in R$ is the feature attribution for feature j. To ensure effective explanations, explanation models often employ simplified inputs, denoted as x', which are mapped to the original inputs x through a mapping function $x = h_x(x')$. It is important to note that while x' may contain less information than x, the mapping function $h_x(x')$ is specific to the current input x, ensuring the approximation $g(z') \approx f(h_x(z'))$ holds true when $z' \approx x'$. Therefore, even with potentially reduced information in x', the mapping h_x guarantees that $h_x(x') = x$.

In their paper, Lundberg and Lee (2017) highlight an intriguing characteristic of the class of additive feature attribution methods. They demonstrate that within this class, there exists a single unique solution that possesses three desirable properties, which were previously unfamiliar to other additive feature attribution methods. These properties closely align with the well-known principles of classical Shapley value estimation methods. The first of these properties is referred to as local accuracy. Local accuracy necessitates that when approximating the original model f for a specific input x, the explanation model should, at the very least, produce an output that matches f when provided with the simplified input x'. It is important to note that x' corresponds to the original input x.

$$f(x) = g(x') = \emptyset_0 + \sum_{i=1}^{M} \emptyset_i x'_i \tag{3.15}$$

Another important property of the additive feature attribution methods is the concept of missingness. When the simplified inputs represent the presence or absence of features, the property of missingness ensures that features missing in the original input have no impact on the attribution. In other words, missingness dictates that a missing feature receives an attribution value of zero. In simplified feature, where x'_j represents the simplified feature and a value of 0 indicates the absence of a feature value, all feature values x'_j of the instance being explained should be set to '1'. The presence of a '0' in the notation would signify that the corresponding feature value is missing for the instance of interest.

$$x'_j = 0 \rightarrow \emptyset_j = 0$$

The third property of additive feature attribution methods, identified as consistency by Lundberg and Lee (2017), plays a crucial role in ensuring the reliability and meaningfulness of the attributions. Consistency states that if a model undergoes changes such that the contribution of a specific simplified input (feature value) increases or remains the same, regardless of the other inputs, the attribution assigned to that input should not decrease. In other words, consistency asserts that any increase or maintenance of the marginal contribution of a feature value in the model should correspondingly result in an increase or maintenance of its assigned Shapley value. This property guarantees that the attributions assigned to the feature values align with the changes observed in their contributions, regardless of the interactions and influences of other features.

4 Experimental Design

4.1 Dataset

The data used in this thesis was the real-world data provided by one of telecommunications company based on Indonesia. The dataset specifically encompasses transactional information related to their prepaid (non-contractual) B2C customers. Prepaid customer was selected to be this research subject because they represent the majority of telco customer in Indonesia, and their constituting 95% of the company total customer base (Telkomsel, 2023).

The dataset contains comprehensive monthly transactional record of 200,000 clients across a time span of three years, starting from January 2020 and concluding in December 2022. To ensure a robust and reliable analysis, only customers who were active in January 2020 and had a minimum length of stay (LOS) of three months in January 2020 were considered. Any customer with less than a three-month LOS was excluded because they could be considered as new or unstable customer (Binh et al., 2021). Furthermore, the selected prepaid customer was from the B2C class, ensuring that they are an individual customer who voluntarily chose to engage with our company's services without any contractual commitments. In addition, to capture diverse consumer behaviours and to offer a balanced view of both active and churned customers, the selected pool consists of 50% customers who remained active throughout the period of observation, while the remaining 50% churned in the last three months of the observation period.

While incorporating demographic data like age, gender, and location can potentially introduce bias, the dataset was curated to minimize such issues. For instance, the age distribution within the dataset aligns with Indonesia's demographic structure (BPS Indonesia, 2022). It could be seen at Figure 11 that the adults (ages 25-64) dominate the dataset at 67.1%, followed by seniors (age > 64) at 18.8%, and young individuals (age < 25) at 14.1%. This distribution minimizes age-related biases and enables a better understanding of service consumption patterns across different age groups. Gender distribution showcases a balanced representation, with males and females almost equally represented. Regarding location, while all regions are well-represented, regions with capital cities displayed a slightly elevated churn rate due to heightened competition, reflecting typical urban customer behavior rather than a dataset bias.

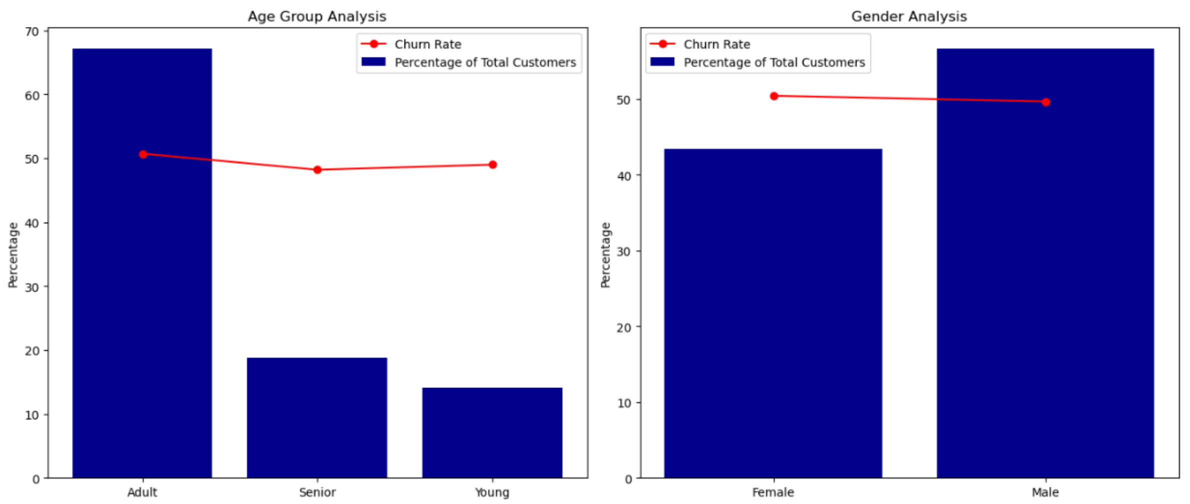


Figure 11 Age group and gender distribution in the dataset

One key aspect of the dataset is the granularity of revenue and usage data. The company offers a range of services, including voice, SMS, broadband, digital, and international roaming. For each of these services, the dataset captures individual revenue and usage behaviour, which is then aggregated on a monthly basis. The usage data capture the total days, total transactions, total duration and volume of services used. Furthermore, it offers insights into customers' online behaviour, categorized according to their intensity of usage and the types of websites or online activities they engage in. As for the demographic data, the dataset includes age, gender, and location of the customers, offering valuable insights into the user base of the company. The dataset also incorporates the details of device used by the customers, providing an insight into the potential influence of device type on customer behaviour and probability to churn. Collectively, these 141 features provide a comprehensive perspective on the customer base, encompassing their behaviour patterns and their financial contributions to the company. Table 5 provides a comprehensive overview of the descriptive statistics for the sample features in the dataset, offering valuable insights into the distribution characteristics of each feature. For a more comprehensive understanding, Appendix 2 contains additional visual representations, including histograms and charts that further illustrate the distributions of the features and also the detailed description of feature name.

Table 5 Descriptive statistics of sample features

	Point	Total revenue	Avg 3m	Total recharge	Avg 3m std	Total Revenue std	Total recharge std
mean	84	79,499	79,410	78,081	32,899	47,488	52,597
std	249	121,963	121,079	132,729	79,412	103,507	106,853

min	-	-	-	-	-	-	-
25%	9	26,490	26,447	25,000	12,917	20,651	22,664
50%	35	57,067	56,978	54,000	23,369	36,172	39,322
75%	101	104,840	104,417	100,000	39,796	58,333	62,641
max	41,299	20,319,820	20,071,050	17,581,750	16,266,500	22,733,530	20,412,410

In addition, the dataset demonstrates a notable level of reliability and validity, due to the automated mechanism employed by the company's system for capturing revenue, usage, and recharge behavior, as well as the comprehensive customer's device data. The automation process effectively mitigates the likelihood of human error, thereby able to maintain the reliable data quality throughout the entire customer base. Furthermore, in terms of privacy preservation and adherence to research ethics, the dataset employs a secure approach by hashing the main identifier of customers, namely the MSISDN or phone number. This approach ensures the anonymity and confidentiality of customer information, highlighting the commitment to maintaining ethical standards in data analysis and safeguarding customer privacy.

4.2 Method to Calculate RAR

In this research, the formula used to calculate risk-adjusted CLV will follow the method calculate CLV that introduced by Berger and Nasr (1998). The CLV model defines the CLV as the present value of the future cash flows generated by a customer, taking into account the customer's generated revenue and the associated servicing costs. To incorporate the time value of money, these cash flows are discounted using a suitable discount rate (Equation 3.1). However, since cost data is unavailable for this research, only the customer's generated revenue is utilized in the calculation Therefore, the metric computed in this study is referred to as RAR instead of RALTV. To calculate RAR, the proposed method builds upon the base model by Berger and Nasr (1998) but incorporates adaptations to the discount rate to account for the risk associated with the customer. Specifically, two customers' risks are considered: the probability of customer churn and the volatility of revenue from each customer that represented by beta value.

The selection of the probability of customer churn is motivated by its frequent occurrence in the telecommunications industry, particularly among prepaid customers who have the flexibility to switch providers at will. Furthermore, this study acknowledges the prevalent role of revenue volatility as a common risk across various industries. To represent this volatility, the study applies the concept of beta value, traditionally used in finance to measure the volatility or systematic risk of a security or portfolio in relation to the market as a whole. This represents a novel approach, suggesting that the method for calculating RAR using revenue volatility, encapsulated by the beta value of customer's revenue contribution, could be valuable and potentially applicable across different sectors. Therefore, by incorporating these two aspects of risk, the churn probability and the beta value of revenue volatility, this methodology for calculating RAR provides not only comprehensive data but also allows for meaningful comparisons across different experiments or industries.

4.2.1 Incorporating the Probability of Customer Churn

The first customers' risk that will be used in the RAR calculation is the probability of churn from customer. The integration the risk of probability of churn into discounted rate formula will follow the approach proposed by Ryals & Knox (2005). Ryals & Knox's model uses two criteria to calculate the discount rate, the risk of customers relative to the market ($\frac{R_c}{R_m}$) and the WACC. They define the customers' risks as the weighted customer credit rating for individual customers, with its average for the entire portfolio representing the market risk (Equation 3.5). In this study, the model was adapted by replacing the customers' risk criterion ($\frac{R_c}{R_m}$) with the individual customer's probability of churn (C_c) divided by the expected probability of churn for the entire market (C_m). And the formula for the discount rate while incorporating the probability of customer churn as the risk is:

$$d = \frac{WACC \times C_c}{C_m} \qquad 4.1$$

For this approach, an assumption is made that the dataset used in this study represents the entire market. Hence the expected probability of churn for the entire market is the average probability of churn from all of the customer in the dataset. The estimation of churn probability is conducted using a selected ML model developed following the CRISP-DM framework and the ML pipeline. In this process, a snapshot of the last 1 year of historical data is chosen to predict the churn probability for each customer. The features from the first 9 months of the snapshot data are utilized to predict the churn probability for the subsequent 3 months. The expected probability of churn for the entire dataset is calculated as the average of the churn probabilities for each customer. This approach, which incorporates the individual customer's probability of churn as the risk criterion, will be referred to as 'approach A' in the later part of this thesis. Additionally, this thesis considers three WACC values of the company: 6.58%, 7.8%, and 7.92% (Alpha Spread, 2023). These values represent the minimum, median, and maximum WACC values observed during the dataset period (January 2020 – December 2022). In the later part of the thesis, RAR calculation using minimum WACC value will be referred as Approach A.1, the medium WACC value will be referred as Approach A.2 and maximum WACC value will be referred as Approach A.3. The RAR results obtained using these three different WACC values will be compared using Welch's ANOVA to evaluate the statistical difference between them.

4.2.2 Incorporating the Volatility of Revenue from Customer into the Discount Rate

The volatility of customer revenue will be quantified using the Beta value. The calculation of the Beta value in this thesis follows the CAPM theory as adopted by Dhar & Glazer (2003), Buhl & Heinrich (2008), and Machado & Karray (2022) in Equation 3.6. Additionally, the discount rate is computed in Equation 3.7. However, the formula is adjusted by replacing the 'returns' for the customer and market with the 'revenues' from the customer and the average revenue from the market. Consequently, the formula to calculate the discount rate when incorporating the volatility of customer revenue becomes:

$$d = \psi_m \, x \, \frac{cov(\varphi_{ct}, \varphi_{mt})}{var(\varphi_{mt})} \qquad 4.2$$

where φ_{ct} *and* φ_{mt} are the revenue for customer c and market m, respectively, at a given time period t and ψ_m represents the expected rate of revenue of the market. In this approach, which will be referred to as "approach B" in this thesis, the market is assumed to represent the entire customer base of the company, providing a comprehensive representation of the revenue dynamics.

To calculate the Beta value, the volatility of revenue from each customer is assessed on a monthly basis. This enables the consideration of variability and fluctuations in the customer's revenue contributions, ensuring that the RAR calculation is sensitive to changes in their revenue patterns. By incorporating the customer revenue volatility, the RAR assessment becomes more comprehensive, accurately capturing the dynamic nature of revenue generation. This enhanced approach provides a robust foundation for evaluating the financial impact of various risks on customer revenue, enabling informed decision-making and resource allocation.

4.3 Experimental Setup And Machine Learning Development Pipeline

The experimental setup of the thesis involves the careful selection of experiments aimed at estimating the RAR value for each customer within the dataset and develop a ML model to predict them. To achieve this objective, firstly, the revenue sources and different risk sources are extracted from the dataset. Subsequently, a method will be proposed to calculate the RAR, tailored to the specific risk sources identified. This method will provide a comprehensive framework for assessing and quantifying the financial impact of various risks on customer revenue. Finally, the obtained results will undergo rigorous statistical tests, including the application of analysis of variance (ANOVA), to compare and evaluate the distinctiveness of the proposed models. These statistical tests will ascertain whether the generated metric exhibits statistically significant differences, thereby substantiating its potential for further exploration.

The RAR result obtained from the previous step serves as the target value for the subsequent development of the ML model. The ML model was developed primarily using Python as the programming language. Figure 12 illustrates the complete ML pipeline employed during the ML model development process. The pipeline was constructed following the CRISP-DM methodology, encompassing a series of key steps and components crucial for effective model development. These steps include data pre-processing, feature selection or engineering, model training, and model evaluation. Each step within the pipeline contributes to the overall process of creating a robust and accurate ML model for predicting RAR. By adhering to the ML pipeline and utilizing the CRISP-DM methodology, the model development process ensures a systematic and comprehensive approach, resulting in a reliable model for RAR prediction.

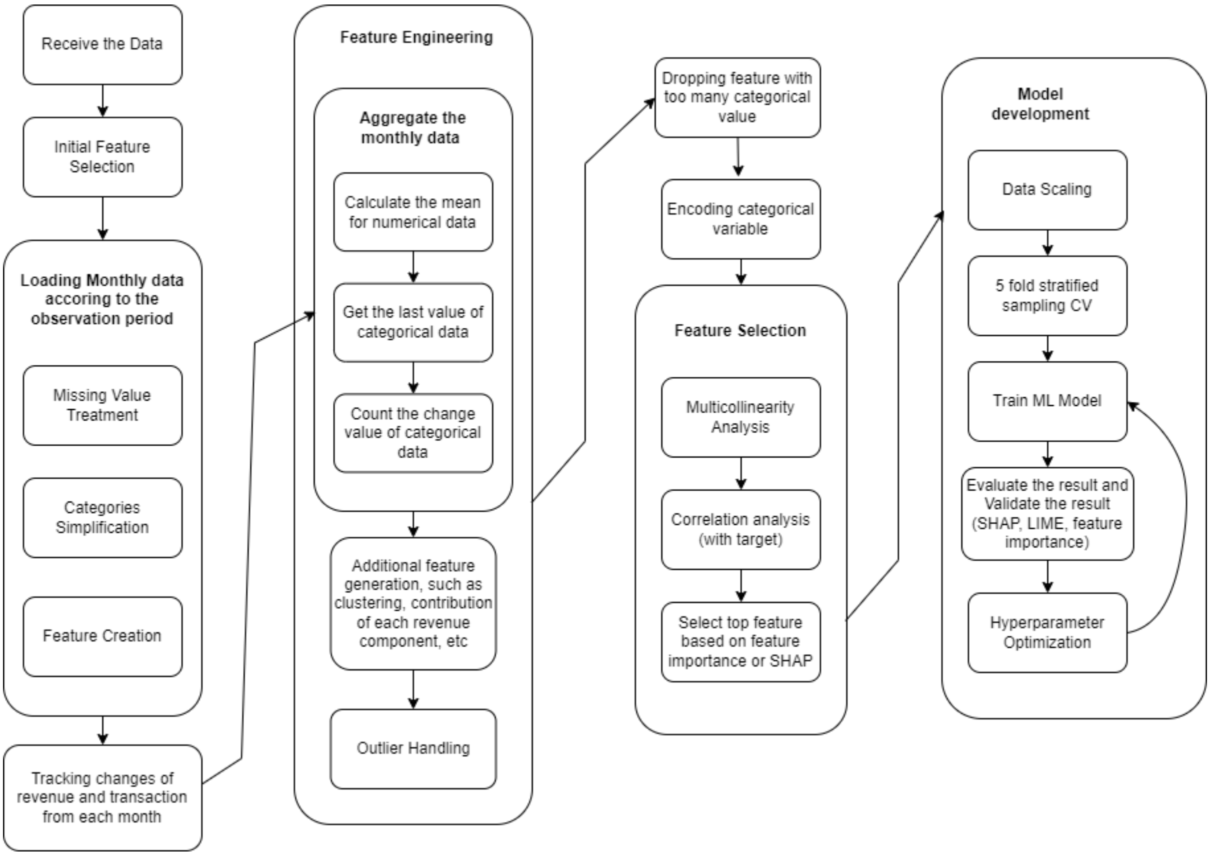


Figure 12 Machine Learning Model Development Pipeline

4.4 Data Preprocessing

An extensive data pre-processing strategy was implemented to guarantee the integrity and dependability of the subsequent analysis. The treatment was customized according to each variable's specific attributes to ensure that the data preprocessing steps were both appropriate and effective.

Initial Feature Selection and data treatment: The first step in data preprocessing involved selecting the main features to be included in the thesis. This process was necessary as the initial dataset received from the company contained various data points that were unrelated to the thesis or overly specific, increasing the data dimensionality without contributing significantly to the thesis objectives. For instance, data related to postpaid customers, which were not relevant to the study focused on prepaid customers, were excluded. Similarly, specific datasets such as the division of SMS into on-net and off-net categories were simplified by including only one value since the other value could be derived by subtracting the overall SMS count from the on-net category. After the initial feature selection, the next step involved addressing inconsistencies in data format and column names. Since the data spanned a three-year period, inconsistencies in column names were bound to occur. To ensure consistency and compatibility, efforts were made to standardize the data format and resolve any discrepancies in column names. This process involved mapping and renaming columns to ensure uniformity throughout the dataset.

Missing Value Treatment: The handling of missing values varied depending on the specific characteristics of each variable. For variables associated with revenue, usage, or recharge, the missing values were replaced with zeroes. It is safe to be done because these values were system-generated and a missing entry implied an absence of data during the observational period. Other variables with missing data were treated differently. For specific columns with rules governing missing value treatment, such as internet category based on the first application or website categories or the customer segmentation according to the average revenue, the rules were strictly followed for the missing value imputation. For several missing value with negligible row of data, the row simply dropped. For categorical variables with missing data, a new category, "No_data", was created, or the mode of the data was used.

Outlier Detection and Treatment: Z-score analysis was applied for outlier detection in each feature. This method is a commonly used technique for identifying outliers by determining if a data point is a certain number of standard deviations away from the mean (Iglewicz & Hoaglin, 1993). Following the identification of outliers, a capping method was used to manage them. This involves limiting the maximum and minimum values for a variable to reduce the impact of extreme values. Furthermore, alternative outlier treatment methods such as Winsorizing and Binning were explored in this thesis. However, the results from both methods were similar and even lower compared to the capping method. As a result, the capping approach was adopted for outlier treatment.

For the RAR prediction task, rows with RAR values identified as outliers were excluded from the analysis. This decision was based on the assumption that these outliers might represent unique situations that could potentially distort the normal image and affect the predictive model. By excluding these outliers, the subsequent analysis focused on more representative data, enhancing the reliability and accuracy of the RAR predictions.

Category Simplification: To handle sparse categories and reduce the noise in the data, several values within some columns were grouped into a single category labeled 'others'. This process helps to ensure a more manageable and robust data set minimizing overfitting due to rare categories.

Encoding Categorical Variables: To prepare the dataset for ML algorithms, categorical variables were transformed via encoding techniques. Ordinal Encoding was utilized for features with binary outcomes, such as gender, while One-Hot Encoding (OHE) was applied to categorical variables with more than two possible values. This encoding approach ensures the algorithms can effectively interpret and utilize these features, thereby enhancing the accuracy and robustness of the resulting analysis.

4.5 Feature Engineering

Feature Engineering plays a pivotal role in ML model development as it will impact the predictive power of the model. On this thesis, this step is divided into two different parts, namely feature extraction and feature selection. Feature extraction process consisted of several steps aimed at generating new variables that could provide additional, meaningful information to the predictive models. And after the new features is extracted, we can proceed into the next step, feature selection. Feature selection aims to identify the most relevant features from the extracted set. This step helps to reduce dimensionality, remove noisy or irrelevant features, and focus on the subset of features that have the most impact on the target variable. Feature selection methods can be applied to the extracted features to identify the subset of features that best contribute to the predictive power of the model.

The Subsection 4.5.1 describes the features extraction process, where it elaborates in the different new features extracted from the existing features and the rationale behind it. In the end of features extraction process, the total features that included in the features selection process amount to 265 features. And the subsequent subsection describes the features selection process.

4.5.1 Feature Extraction

The feature extraction step conducted in this thesis are:

1. **Binning/Grouping**: To enhance the data representation and improve interpretability, specific variables, including age and region were subjected to binning or grouping. Age was categorized into three groups: young, adults, and seniors. This categorization allows for a more comprehensive analysis of the different age segments and their impact on the desired outcomes. Additionally, the region variable was binned into four distinct

tiers based on factors such as economic power and population size. This categorization provides a more nuanced understanding of regional differences and enables the exploration of their influence on the study variables.

2. **Days since Last Transaction:** As models often struggle to interpret datetime data directly, the difference between the last transaction (either service usage or recharge) and the recording date was calculated. By converting these datetime values into a more digestible integer format, the model can more effectively utilize these features.
3. **Remaining Active Period:** The remaining active period of a customer was calculated by subtracting the recording date from the grace period. Similar to the previous point, converting datetime data into integer values makes it easier for models to interpret the data.
4. **Service Usage:** New columns were generated to track the frequency of customers purchasing voice, broadband, or SMS packages. This feature was engineered to capture customer price sensitivity. Regular customers who frequently purchase packages for their most-used services could be considered price-sensitive. On the other hand, regular customers who seldom purchase packages, even for their most-used services, might be less price-sensitive.
5. **Change in Recharge and Revenue:** To capture the volatility in a customer's recharge and revenue patterns, the percentage change and standard deviation of both total revenue and individual service revenues were computed. This feature could offer insights into the stability of a customer's financial contribution and potentially signal impending churn.
6. **Revenue Contribution:** The contribution of each revenue component to the total revenue was calculated. This could highlight the relative importance of each service in generating revenue, might provide insights into the revenue structure and customer's habit.

4.5.2 Feature Selection

Following the creation of new features, the process moved on to the selection of the most relevant features for the predictive model.

1. **Dimensionality Reduction**: High-dimensionality data can potentially overfit the model and introduce noise, thus compromising model performance. Therefore, certain columns that encapsulate high-dimensionality information, such as sub-district data, device data, and most-used application data, were eliminated. These high-

dimensionality features were replaced with more generalized counterparts, such as region data in place of sub-district data and device type data instead of specific device data. This reduction of dimensionality serves to improve model interpretability and reduce computational complexity while preserving essential information for the model.

2. **Multicollinearity Analysis:** To avoid redundancy and to improve the interpretability of the model, multicollinearity analysis was performed on the 265 remaining features. Features that exhibited correlation greater than 85% with any other feature were eliminated to prevent multicollinearity, which can adversely affect the performance and interpretability of some models.
3. **Correlation Analysis:** Following the multicollinearity analysis, an evaluation of the remaining features occurred based on their correlation with the target variable. For the preliminary model training, the choice fell on the top 150 features demonstrating the highest correlation with the target. This selection stemmed from the consideration that data ranked beyond these features already exhibited very low correlation with the target (Appendix 2). Selecting a substantial quantity of 150 features ensures a wide array of influencers available for subsequent SHAP analysis, thus improving the potential for a more robust model. This method ensures that chosen features display a significant relationship with the target, enhancing their potential to contribute meaningfully to the predictive power of the model.
4. **Feature Importance Analysis**: The final selection of features was made based on their importance as determined either from the embedded ML function or from the global importance of SHAP values. This step further refined the features to the top 10-20 that had the most significant impact on the model's performance. Using this feature importance analysis provides a robust method of selecting the most impactful variables and reducing overfitting, improving the generalizability of the predictive model.

4.5.3 Feature Scaling

Data scaling is an essential transformation applied during the data pre-processing stage to address the issue of disparate value ranges across features. When feature values exhibit significant variations, ML models may inadvertently assign more weight to features with higher average values, potentially biasing the model's predictions. To mitigate this bias, feature scaling techniques are employed to normalize the values and ensure fair treatment of each feature.

Two widely used feature scaling methods are min-max scaling and standardization (García et al., 2014). Min-max scaling rescales the feature values to a range between zero and one, effectively mapping the data to a uniform scale. On the other hand, standardization transforms the values to have a mean of zero and a standard deviation of one. While standardization is less sensitive to outliers, it may not be the most suitable choice when working with some ML algorithm (Sefara, 2019). Given that the dataset in this study contains either minimal outliers or has successfully addressed them during the pre-processing steps, the min-max scaling approach has been deemed appropriate for the feature scaling process.

In this thesis, the feature scaling process was carefully implemented within a pipeline framework to mitigate the risk of data leakage during model training. Data leakage refers to a situation where information from the test or validation set inadvertently leaks into the training set, leading to overly optimistic performance metrics and compromised model generalization (Kaufman et al., 2012). By incorporating feature scaling within the pipeline, the scaling transformation is applied to the training data while keeping the scaling parameters isolated from the test or validation data. This ensures that the scaling process is performed solely based on the training data distribution and characteristics. Consequently, when the model is evaluated using unseen data, the feature scaling is applied consistently, reflecting the real-world scenarios and enhancing the model's ability to generalize to new instances.

4.6 Model Development

This section details the construction of two ML models: the customer churn prediction model and the RAR prediction model. To ensure a comprehensive comparison of model performance, four distinct ML algorithms were employed: LR, RF, XGBoost, and Catboost. The selection of these algorithms was driven by their unique characteristics. LR, known for its simplicity and statistical efficiency, provided a baseline model for comparison. On the other hand, RF, XGBoost, and Catboost were chosen for their complexity and ensemble-based approach. RF belongs to the bagging/boostrapping algorithms, while XGBoost and Catboost utilize boosting learning algorithms. By comparing these different learning methods, valuable insights can be gained regarding their performance in the specific use case of this thesis.

Prior to train the model, it is essential to address the data splitting strategy used to ensure reliable and robust model evaluation. In this thesis, a stratified sampling five-fold cross-validation (CV) approach was adopted to split the data prior to training the models. Conventional five-fold CV entails randomly dividing the dataset into five equal-sized folds, with each fold serving as a validation set once while the remaining four folds are allocated for training. However, in this thesis, the stratified sampling technique was employed to ensure the preservation of the class distribution within each fold. By doing so, it was ensured that each fold contained approximately the same proportion of samples from each class as the original dataset. This approach is crucial as it mitigates the risk of biased evaluation and ensures consistent assessment of the model's performance across different folds (Kohavi, 1995). Furthermore, while the available data initially consisted of monthly records, the model development process involved aggregating the data without considering the precise sequence of individual data points. In this context, splitting the data without explicitly accounting for the temporal order was deemed appropriate. This approach enabled the utilization of a larger and more representative dataset for model training and evaluation, enhancing the robustness and generalizability of the results.

The model development process involved multiple steps. After having 5 set of data resulted from the 5-fold stratified CV, then the base model could be trained using the 150 features selected during the feature selection phase. Further refinement of the model was carried out by deriving feature importance from the embedded ML function or the global importance of SHAP values to identify the final features for inclusion in the model. For Churn prediction model, Principal Component Analysis (PCA) was also considered in the dimensionality reduction and feature extraction method. However, the model performance result using PCA was still slightly worse compared to selection features using features importance, hence the final feature selection method was using feature importance. However, in the RAR prediction model, the method of using PCA was not considered to be implemented. It is because one of the final goals is also to understand the contribution of individual features to the RAR prediction model's performance, hence using PCA for feature extraction would not be appropriate. The final number of selected features was decided from a small experiment to decide the most optimal number of features.

Subsequently, hyperparameter optimization was applied to enhance the performance of the model. The hyperparameter was optimized using the combination of random search and grid search. Random search was used as initial exploration of set hyperparameter to narrow down to a promising region, then grid search was employed to fine-tune the parameters within that region. This combination can be a good strategy to balance computational efficiency and optimization performance. The hyperparameter tuning process was carried out using both CPU (LR and RF) and GPU (XGB and Catboost) resources, leveraging their respective capabilities.

This allowed for efficient exploration and optimization of the hyperparameter space. The final set of hyperparameters used in this thesis is presented in Table 6.

Table 6 Hyperparameter used for each model

Model	Algorithm	Hyperparameter
Churn	LR	max_iter = 1000, penalty = 'l1', C = 0.01, solver = 'liblinear'
	RF	n_estimators = 500, min_samples_split = 10, min_samples_leaf = 2, max_features = 'auto', max_depth = None,
	XGBoost	subsample = 0.8, n_estimators = 500, max_depth = 6, learning_rate = 0.05, gamma = 0.1, colsample_bytree = 0.8,
	Catboost	n_estimators = 1000, learning_rate = 0.05, depth = 8,
RAR – Beta Value Approach	RF	bootstrap = True, max_depth = None, max_features = auto, min_samples_leaf = 3, min_samples_split = 2, n_estimators = 500
	XGBoost	colsample_bytree = 1.0, gamma = 0, learning_rate = 0.05, max_depth = 9, n_estimators = 1000, subsample = 1
	Catboost	Depth = 9, learning_rate = 0.15, loss_function = MAE, n_estimators = 800,
RAR – WACC and Churn Approach	RF	Bootstrap = True, max_depth = None, max_features = auto, min_samples_leaf = 2, min_samples_split = 2, n_estimators = 300
	XGBoost	colsample_bytree = 1.0, gamma = 0, learning_rate = 0.05, max_depth = 9, n_estimators = 900, subsample = 0.8

Catboost	n_estimators = 1100, depth = 9, learning_rate = 0.08, loss_function = MAE,

4.7 Model Evaluation and Validation

The ML models developed in this thesis can be categorized into two types: the churn prediction model, which is a classification model, and the RAR prediction model, which is a regression model. Each model requires specific evaluation metrics tailored to its respective task.

For the churn prediction model, the evaluation focuses on two primary metrics: accuracy and the F1 score. Accuracy measures the overall correctness of the model by calculating the proportion of correct predictions out of all predictions made. However, because the churn model involves identifying potential churners, precision and recall values are also crucial. Therefore, the F1 score, which is the harmonic mean of precision and recall, is employed as an additional measure. A high F1 score indicates a well-balanced performance between precision (the ratio of true positive results to the sum of true positive and false positive results) and recall (the ratio of true positive results to the sum of true positive and false negative results), providing a comprehensive understanding of the model's performance in churn prediction.

On the other hand, the evaluation of the RAR prediction model utilizes MAPE and RMSE as the main metric. MAPE represents the average of the absolute percentage errors, allowing for an intuitive comprehension of error rates in percentage terms. However, MAPE treats all data points equally and may not adequately capture the impact of high variance in a small subset of data. In contrast, RMSE provides information on the average magnitude of prediction errors, placing higher emphasis on large errors due to its squaring operation. The other metric also evaluated in the model, such as: MAE and R^2.

To ascertain the robustness of the final model, various splitting strategies were examined to assess the model's performance under different conditions. These strategies utilized different split ratios ranging from 0.5 to 0.9. In addition, the final model's performance was evaluated using 5-fold CV and group-split CV. Given the regression nature of the RAR model, a five-fold CV (without stratified sampling) suffices. Furthermore, group-split CV was applied to supplement the evaluation process, with the dataset divided according to region_tier.

To further assess model performance, a learning curve analysis was conducted by training the model with varying amounts of training data.

Furthermore, the model validation in this thesis will employ XAI techniques, including Feature Importance, SHAP and LIME These methods enable a comprehensive understanding of the specific contributions of features to the model predictions, which is crucial in this context. Feature Importance and SHAP provide insights into the variables with the greatest influence on predictions from a global perspective. Furthermore, SHAP and LIME provide granular, localized insights from individual data points, providing a deeper understanding of the model's behavior. For local explanations in approach A, two customers were evaluated: sample X, representing customers who stayed, and sample Y, representing those who churned. Similarly, in approach B, two distinct customers were used: sample X, which denotes customers with a negative beta value, and sample Z, which symbolizes those with a positive beta value. It's worth noting that sample X in both approaches represents the same customer. By utilizing these XAI techniques, the correlation between the most impactful features identified by the models and the elements involved in the RAR calculation could be established, validating the model from both performance and theoretical perspectives.

5 Results and Discussions

5.1 Probability of Customer Churn Model

The approach A of this thesis use the probability of customer churn model, hence prior the RAR calculation, the churn model will be developed first. As mentioned in the previous chapters, the churn model will be developed following the ML pipeline and using four different algorithms (LR, RF, XGB and Catboost). Figure 13 illustrates the average results of a 5-fold cross-validation for evaluating the model's performance across different numbers of features. The average accuracy and average F1 score were computed to assess the impact of feature selection on the model's predictive capabilities.

The findings reveal a consistent pattern of improvement in both average accuracy and average F1 score as the number of features increased. From 5 to 9 features, there were incremental improvements observed, with average accuracy ranging from 0.835 to 0.847 and average F1 score ranging from 0.821 to 0.844. Afterwards, the model improvement is minimal and had notable performance improvements at 18 features, reaching a value of accuracy 0.852, and average F1 score, reaching a value of 0.848. Hence, the number of features selected on the model development is 18 features.

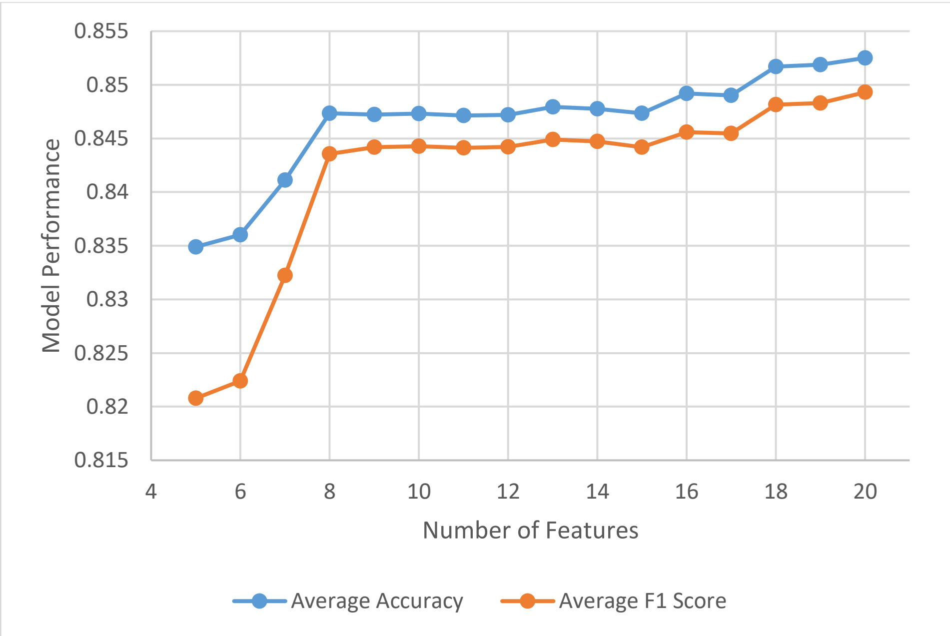


Figure 13 Total feature vs model performance for probability of churn model (5-fold CV)

5.1.1 Model Performance Evaluation Metric

Table 7 presents the performance results of each model. The logistic regression model, known for its simplicity, allowed for direct optimization of hyperparameters using grid search. When evaluating feature selection, the analysis revealed negligible differences in performance between utilizing all available features and selecting only the top 18, accuracy and ROC-AUC scores reduced by a mere 0.01. Interestingly, despite the grid search optimization, no significant improvements in the model's overall performance were observed.

Table 7 Probability of Customer Churn: Model Performance (5-fold CV)

Model & Model Development Stage	Accuracy	Recall	Precision	F1 Score	ROC-AUC
Logistic Regression					
Initial model and all features	0.80	0.77	0.81	0.79	0.80
Feature selection	0.79	0.70	0.86	0.77	0.79
Grid search hyperparameter tuning	0.79	0.70	0.86	0.77	0.79
Catboost					
Initial model and all features	0.853	0.832	0.870	0.850	0.853
Feature selection	0.852	0.828	0.869	0.848	0.852
Random search hyperparameter tuning	0.854	0.847	0.859	0.853	0.854
Grid search hyperparameter tuning	0.856	0.844	0.865	0.854	0.856
XGBoost					
Initial model and all features	0.890	0.886	0.894	0.890	0.890
Feature selection	0.889	0.884	0.893	0.889	0.889
Random search hyperparameter tuning	0.888	0.887	0.890	0.888	0.888
Grid search hyperparameter tuning	**0.892**	**0.890**	**0.893**	**0.892**	**0.892**
Random Forest					
Initial model and all features	0.799	0.722	0.854	0.782	0.799
Feature selection	0.796	0.695	0.871	0.773	0.796
Random search hyperparameter tuning	0.789	0.653	0.896	0.756	0.789
Grid search hyperparameter tuning	0.797	0.694	0.875	0.774	0.797

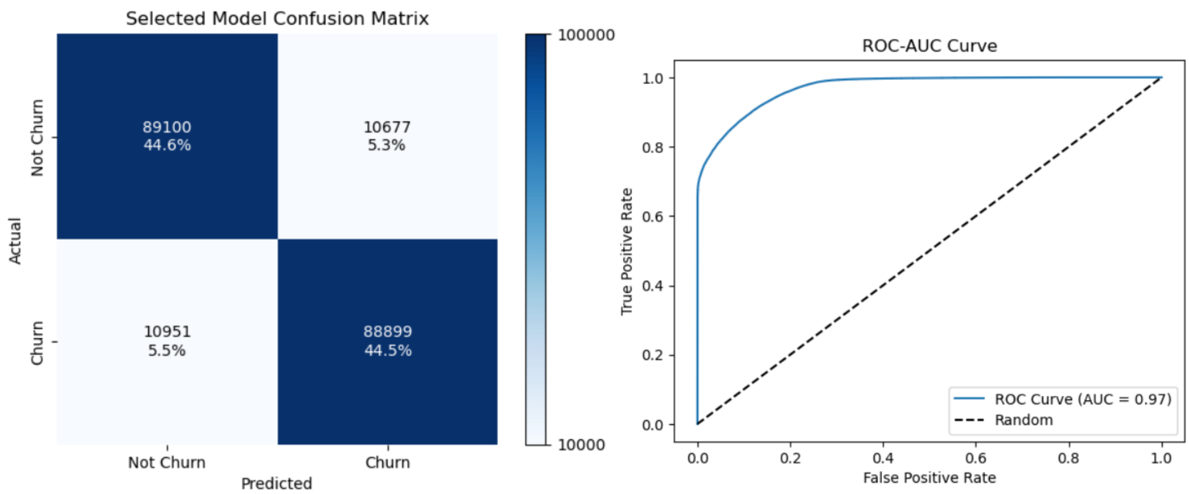
The performance of the Catboost algorithm surpasses that of the logistic regression model significantly, with 0.05 difference in accuracy, recall, F1 Score and ROC-AUC score between them. When considering only the top 18 features in Catboost, the accuracy and F1 score remain highly similar with using all of the features, with only a marginal decrease of 0.001. The hyperparameter optimization process further enhances the model's performance, although the improvements are relatively small. Notably, the final Catboost model exhibits well-balanced recall and precision values, with a minor difference of only 0.02 between them, resulting in a high F1 score.

On the other hand, the performance of XGBoost model outperforms both logistic regression and Catboost models by a significant margin, with the all the metric almost reaching 0.9. Notably, the recall and precision metrics exhibit stability, with minimal differences observed across different stages of model development. Additionally, the grid search hyperparameter optimization process proves effective in enhancing the performance of all metrics in the final model.

The RF algorithm yields results similar to logistic regression, with the final model achieving an accuracy of less than 0.8. Moreover, when selecting only the top 18 features, there is a significant reduction in the recall value, while the precision value shows only slight improvement. Similar to logistic regression, the final Random Forest model performs slightly worse in terms of metrics compared to the initial model using all available features.

5.1.2 The Best Model

Among the evaluated models, the XGBoost model consistently demonstrated superior performance in terms of accuracy, recall, precision, F1 score, and ROC-AUC. Additionally, the grid search hyperparameter optimization further improved the model's performance across all metrics and resulted in a balanced outcome with high accuracy and F1 score. The final model used in this thesis is the XGBoost model, which achieved an accuracy and F1 score of 0.892. Detailed results of the best model, including the confusion matrix and ROC-AUC curve, are presented in Figure 14.



(a) (b)

Figure 14 Customer churn prediction model: (a) Confusion matrix and (b) ROC-AUC curve

To gain insights into the factors driving customer churn, a feature importance analysis using the SHAP value was conducted, and the results are shown in Figure 15. The analysis revealed that the most important feature in predicting the probability of customer churn is the "brand_A" feature, indicating the type of SIM card brand used by the customer. The second and fourth most important features are the standard deviation of recharge transactions and SMS revenue. Interestingly, among the top 18 features, a significant proportion (10 out of 18) are related to the customer's region. These region-related features include "region_tier," which represents the economic power and population of the region, as well as specific regions identified by "region_lacci_xx." Additionally, the feature "region_lacci_nunique" indicates the number of different regions where the customer resided during the observation period.

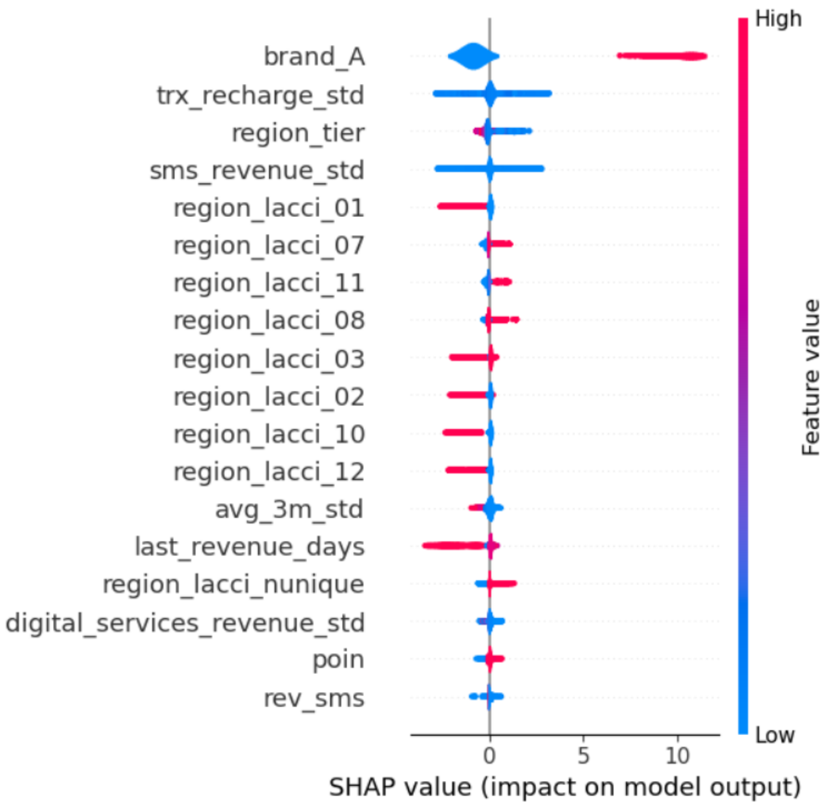


Figure 15 Customer churn prediction model: SHAP Value

5.2 RAR Calculation

5.2.1 Approach A

The approach A for calculating RAR incorporates the probability of churn obtained from the previous model and the WACC values observed during the observation period. Three different WACC values are considered: the minimum WACC value (approach A.1), the median WACC value (approach A.2), and the maximum WACC value (approach A.3). The statistical distribution of the actual total revenue and the calculated RAR values is presented in Table 8 and the histogram of calculated RAR values is presented in Figure 16. To ensure a reliable distribution, any customers with outlier RAR values were excluded from the dataset, resulting in a remaining sample size of 197,692 customers. The distribution in Table 8 also showed that the RAR values are consistently lower than the total revenue, as expected. Additionally, as the WACC value increases, the RAR values decrease. This relationship is anticipated because the discount rate used to calculate RAR is highly correlated with the WACC value (as defined in Equation 4.1). Consequently, a higher WACC value leads to a higher discount rate, resulting in lower RAR values.

Table 8 Approach A: Descriptive statistic of Calculated RAR Result

	Total	RAR value		
	Revenue	WACC min	WACC mid	WACC max
count	197,692	197,692	197,692	197,692
mean	2,683,918	2,255,263	2,195,887	2,190,303
std	2,144,940	1,855,696	1,823,163	1,820,201
min	2,550	1,697	1,588	1,577
25%	1,110,838	903,632	870,937	867,880
50%	2,113,714	1,738,899	1,682,397	1,676,812
75%	3,580,973	3,024,572	2,948,474	2,940,945
max	17,724,060	10,821,540	10,821,000	10,820,950

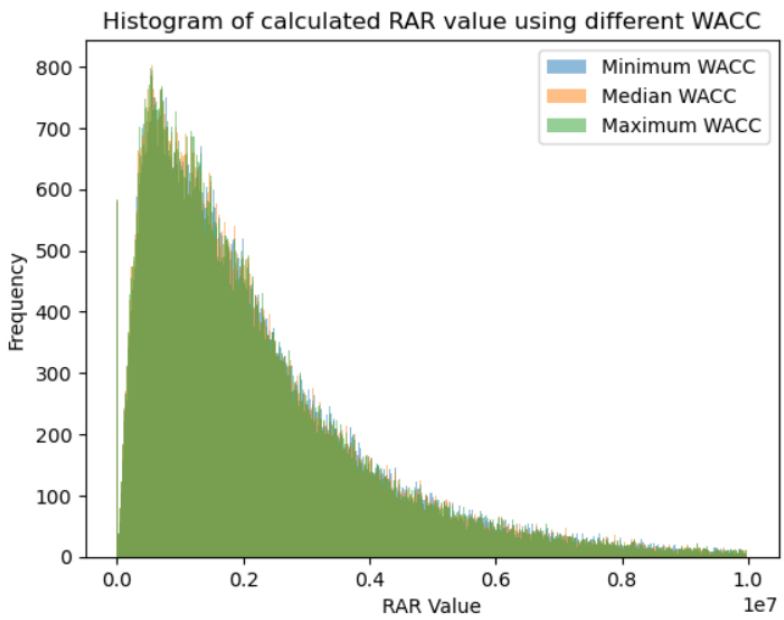


Figure 16 Approach A: Distributed RAR value

Figure 17 displays a scatter plot illustrating the relationship between total revenue (x-axis) and the calculated RAR value (y-axis), with each dot representing a customer in the dataset. The color coding of each dot represents the impact of probability of churn. The plot reveals that higher probabilities of churn correspond to lower RAR values. This observation aligns with the expectations outlined in Equation 4.1, highlighting the influence of both the WACC and the probability of churn on the RAR calculation and their significance in the financial assessment of customer revenue. Furthermore, the scatter plot demonstrates that customers with high probabilities of churn are distributed across a wide range of total revenue values. This finding is further emphasized in the box plot (Figure 18), which depicts the distribution of RAR values across different bins of probability of churn. Each bin exhibit similar distribution of RAR values, indicating that the impact of churn probability on RAR is consistent across various segments of customers.

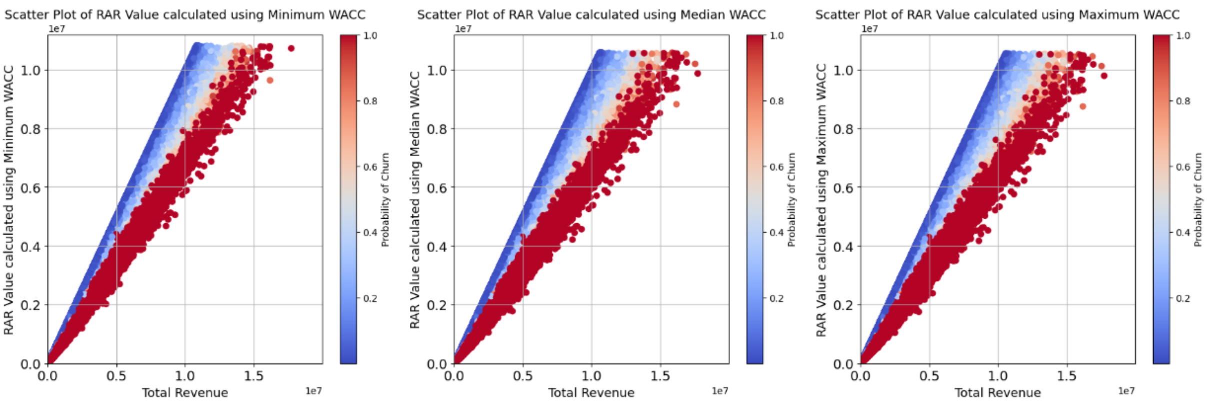


Figure 17 Approach A: Scatter plot of Total Revenue vs RAR Value

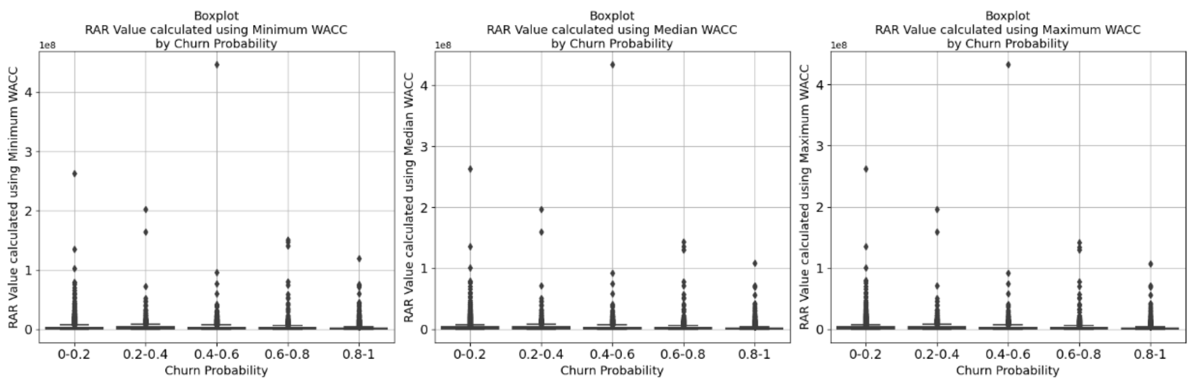


Figure 18 Approach A: Boxplot of RAR value to Probability of Churn

The impact of probability of customer churn to RAR value is more prominent in the Figure 19. As the probability of churn decreases, the ratio of RAR to total revenue increases. These visualizations provide valuable insights into the relationship between total revenue, probability of churn, and RAR, offering a comprehensive understanding of how these factors interplay in the assessment of customer revenue.

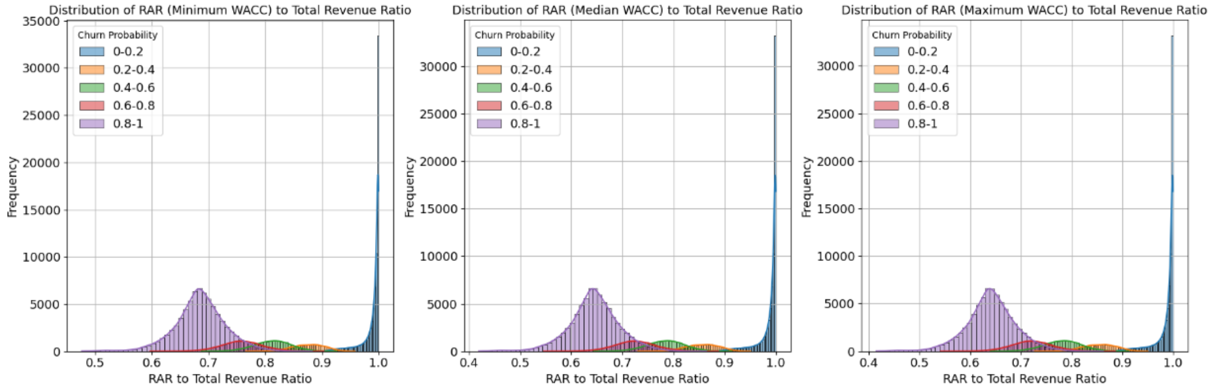


Figure 19 Approach A: Distribution of RAR value to total Revenue Ratio and the Probability of Churn

5.2.1.1 Statistical Analysis to Compare RAR Value

According to the findings presented in Table 8 and Figure 16, the calculated RAR values using three different WACC values exhibit similar results. A statistical test is needed to determine the presence of statistically significant differences in the calculated RAR values across the three methods within approach A. The ANOVA is an appropriate test for comparing the means of multiple groups. However, the assumption of homoscedasticity (equal variances) required for traditional ANOVA is not met in the data.

To assess the assumption of homogeneity of variance, both Levene's and Bartlett's tests were conducted. The results of these tests indicated p-values of 9.36e-08 and 1.99e-41, respectively, both significantly lower than the standard significance level of 0.05. This suggests evidence against the assumption of equal variances in the data. As a result of the detected heteroscedasticity, a Welch's ANOVA was performed. Welch's ANOVA does not assume equal variances and is therefore more suitable for our data. The Welch's ANOVA yielded a p-value of 2.85e-21, which is significantly below the 0.05 threshold. This indicates that there are statistically significant differences in the RAR values obtained from the three methods within approach A.

5.2.2 Approach B

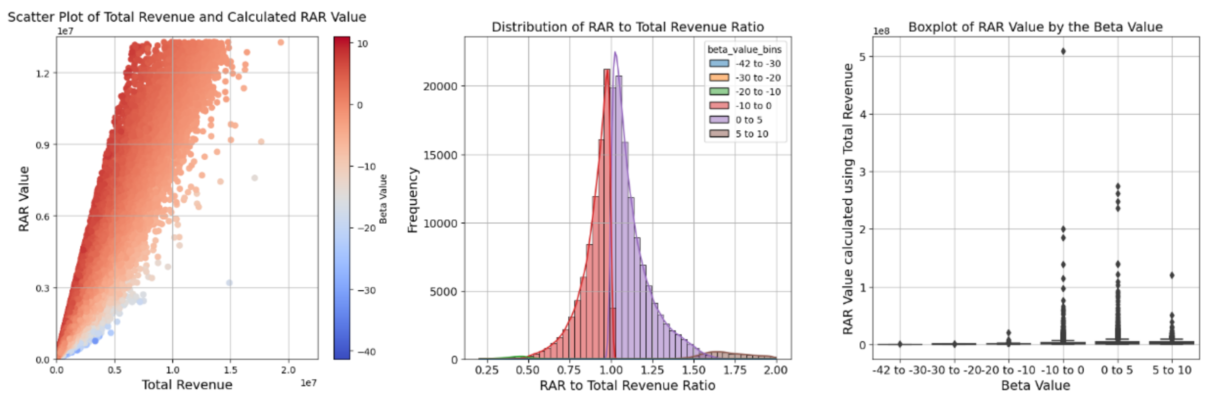
In the results for Approach B, there's a detailed explanation of how RAR is calculated, based on the volatility of revenue indicated by the Beta Value. Table 9 displays the statistical distribution of these RAR values, highlighting a significant difference from Approach A. In

Approach B, the RAR values can be higher than the actual total revenue. This situation is expected as the discount rate used in this approach could have both positive or negative value. The discount rate is the result of the expected return of all customers in the company multiplied by the Beta value, as depicted in Equation 4.2. The expected return from all customers in the company is -0.013, hence negative beta value will lead to positive discount rate and positive beta value lead to negative discount rate and produced the higher RAR value compared to the actual total revenue.

Table 9 Approach B: Descriptive statistic of Calculated RAR Result

	Total Revenue	RAR Value
count	194,464	194,464
mean	2,702,213	2,809,040
std	2,164,500	2,279,828
min	2,550	1,652
25%	1,117,476	1,131,818
50%	2,125,282	2,185,314
75%	3,601,637	3,765,111
max	23,360,000	13,277,820

Figure 20 provides insights into the distribution of Beta values, the observed beta value is ranging from -42 until 10. However, on scatter plot that very small amount of customer with beta value less than -10 and showing that the majority of customers have Beta values ranging from -10 to 5. The histogram plot emphasizes the relationship between Beta values and the RAR-to-total-revenue ratio, demonstrating that higher Beta values correspond to higher ratios. Additionally, the boxplot further illustrates that higher Beta values are associated with higher calculated RAR values.



(a) *(b)* *(c)*

Figure 20 Approach B: (a) scatter plot, (b) histogram, (c) boxplot of RAR value to total revenue and beta value

5.3 RAR Prediction Model: Approach A.1 (Minimum WACC value)

Figure 21 shows the model performance, as measured by MAPE and RMSE, for each number of features used in the model. From the Figure 20, it could be seen that as the number of features increased from 5 to 9, the average MAPE exhibited unstable performance, with minor fluctuations observed. However, from 9 features onwards, there was a consistent improvement in the average MAPE, indicating enhanced predictive accuracy. Notably, the improvement became more substantial as the number of features increased from 9 to 17, reaching its lowest point at 23.45% for 17 features. In contrast, the average RMSE demonstrated a continuous improvement as the number of features increased. This suggests that incorporating additional features into the model resulted in a better fit to the observed data, thereby reducing the overall prediction error. The average RMSE consistently decreased from 591,493 for 5 features to 557,624 for 17 features.

The final selected number of features was determined to be 17, as this configuration offered the most effective balance between predictive accuracy and model complexity. While the average RMSE continued to improve beyond 17 features, the magnitude of improvement became marginal. This indicates that incorporating additional features beyond 17 did not lead to substantial gains in reducing the prediction error. On the other hand, the average MAPE demonstrated a consistent improvement until 17 features were included. This suggests that the selected features effectively captured the underlying patterns and relationships in the data, resulting in enhanced predictive accuracy. Therefore, considering both the marginal improvement in RMSE beyond 17 features and the consistent improvement in MAPE, the decision was made to select 17 features as the optimal configuration for the model.

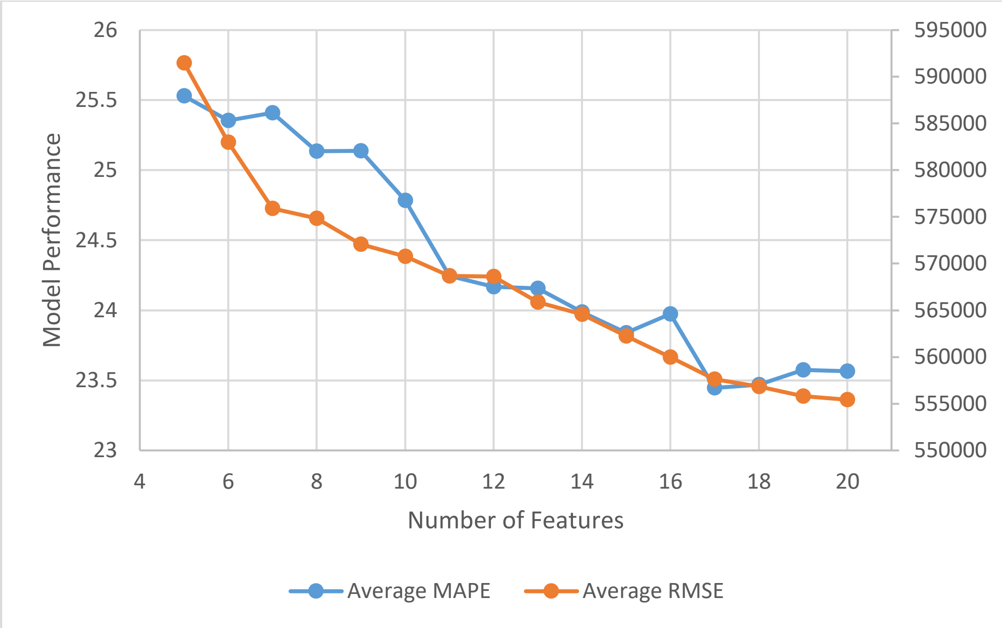


Figure 21 Approach A.1: Number of feature vs model performance (5-fold CV)

5.3.1 Model Performance Evaluation Metric

Table 10 displays the performance metrics of each model during the model development stage. Notably, the Catboost algorithm consistently demonstrated strong performance throughout the modeling process. The initial model achieved an impressive R^2 value of 0.91 and a MAPE of 23.36%. After feature selection, the model's performance experienced minimal degradation, with the R^2 value remaining at 0.91 and the MAPE decreasing slightly to 23.64%. In the final model, significant improvements were observed across all metrics, except for RMSE, which exhibited a slight increase of 10000. The MAPE saw a notable reduction of 4%, indicating enhanced accuracy and precision.

Table 10 Approach A.1: Model performance (5-fold CV)

Model & Model Development Stage	MAPE	MAE	RMSE	R^2
Catboost				
Initial model and all features	23.36	358,254	548,076	0.91
Feature selection	23.64	362,555	555,542	0.91
Random search hyperparameter tuning	19.51	355,476	560,046	0.91
Grid search hyperparameter tuning	**19.38**	**353,781**	**558,315**	**0.91**
XGBoost				
Initial model and all features	23.74	369,528	565,848	0.91
Feature selection	23.66	371,903	569,962	0.91

Random search hyperparameter tuning	22.45	363,472	559,686	0.91
Grid search hyperparameter tuning	21.94	360,894	557,307	0.91
Random Forest				
Initial model and all features	23.13	379,640	584,747	0.9
Feature selection	23.09	381,036	585,703	0.9
Random search hyperparameter tuning	22.58	379,528	584,432	0.9
Grid search hyperparameter tuning	22.56	378,131	582,710	0.9

Similarly, the XGBoost algorithm exhibited strong performance in the initial model, with a MAPE value of 23.74% and an R^2 value of 0.91, which were comparable to the performance of the Catboost algorithm. As the model development progressed, further improvements were observed in all metrics. The MAPE saw a reduction of 2%, indicating enhanced accuracy and precision in the final model. Additionally, the RMSE decreased by 7500, indicating a better fit of the model to the data. The R^2 value remained consistently high at 0.91 throughout the modeling process.

The Random Forest model exhibited a strong initial performance, achieving the best MAPE value among the evaluated models. However, it also had the highest initial RMSE. Throughout the model development process, the improvements were relatively minor. The MAPE showed a marginal improvement of only 0.6%, the RMSE decreased by 3000 units and the R^2 value stayed at 0.9.

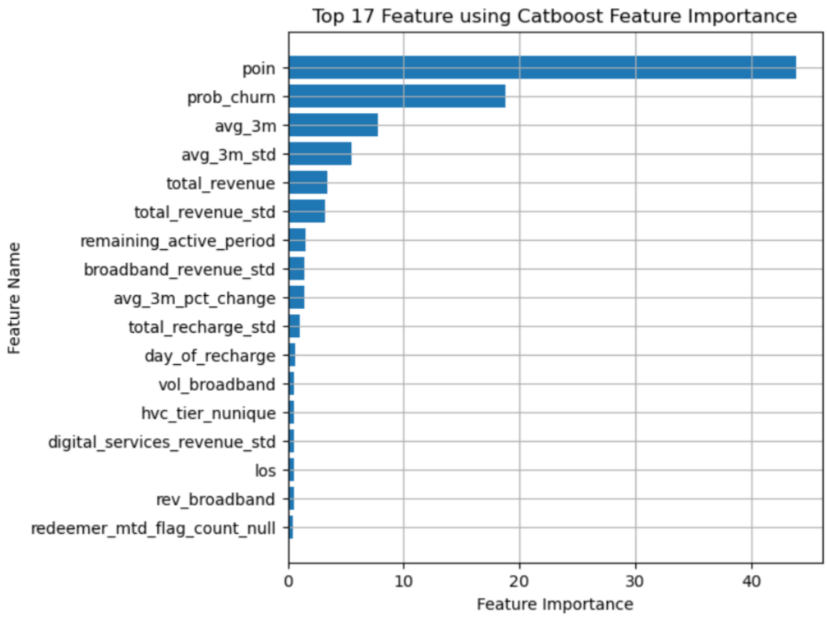
In summary, the Catboost algorithm exhibited the best overall performance, delivering the lowest MAPE in the final model. The XGBoost algorithm also demonstrated strong performance with improvements across all metrics. On the other hand, the Random Forest model showed a strong initial performance but had limited advancements during the model development process. Hence, Catboost was selected as the final model due to its balanced performance in all metrics.

5.3.2 Model Interpretation and Feature Importance Analysis

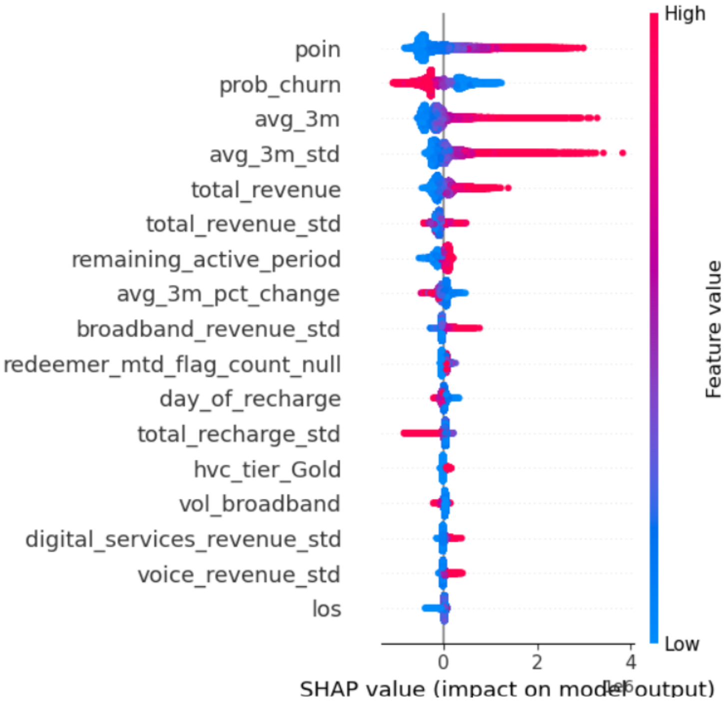
5.3.2.1 Global Explanation

To further evaluate and validate the model, feature importance analysis was conducted using both the feature importance function embedded in the model and the global interpretation of SHAP values. The results are presented on Figure 22 and revealed that the top 9 most important features were consistent across both methods. The most influential feature, as identified by the feature importance method, was "poin," which represents the total loyalty points received by the customer from past transactions. This feature had the highest importance value and exerted a significant impact on the model's predictions. It was found that the importance of "poin" was more than twice that of the second most important feature, which was the probability of churn. However, when considering the SHAP global importance, the dominance of the "poin" feature on the prediction results was not as pronounced as indicated by the feature importance. The SHAP values provided a comprehensive understanding of the feature's impact, specifically illustrating how different feature values contribute to the predicted RAR values. For example, the SHAP values revealed that higher values of "poin" were associated with higher RAR values, suggesting a positive relationship between customer loyalty points and revenue.

The probability of churn was identified as the second most important feature in both the feature importance and SHAP analyses. This emphasizes its significant role in predicting churn behavior. Additionally, the average revenue over the last 3 months ("avg_3m") was ranked as the third most important feature, highlighting the importance of recent revenue trends in understanding customer churn.



(a)

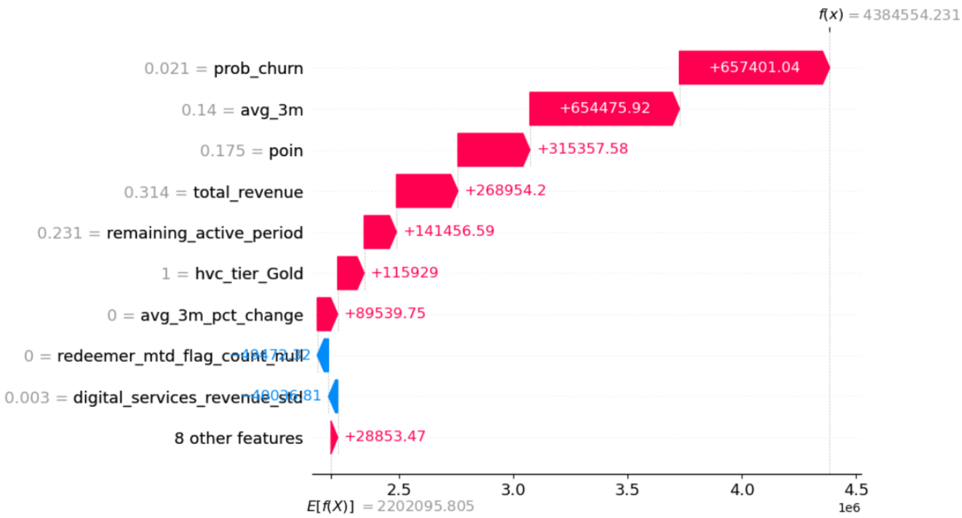


(b)

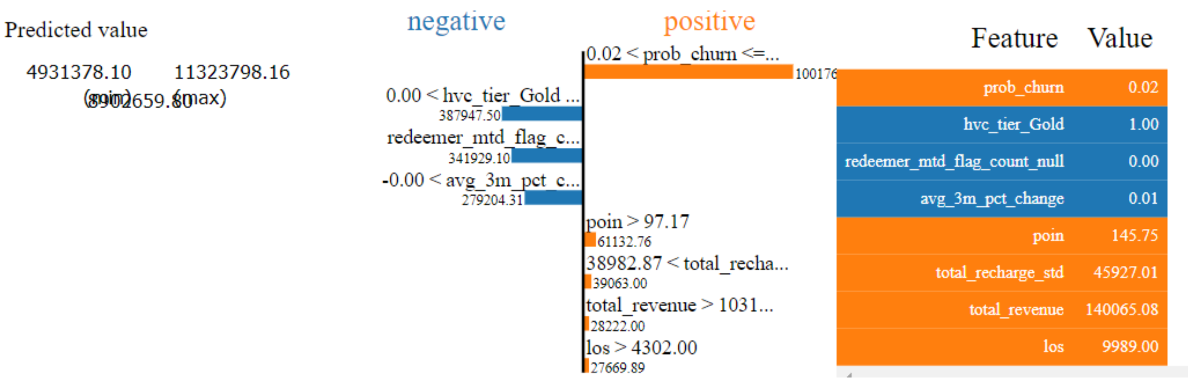
Figure 22 Approach A.1: (a) Feature Importance and (b) SHAP value

5.3.2.2 Local Explanation

To further evaluate the impact of each feature on the prediction value, a more detailed analysis was conducted using SHAP waterfall plots and LIME. These methods provide insights into the contribution of each feature for individual instances, allowing for a more granular understanding of their influence on the predictions. A sample of data that represent stayed customer, sample X, was selected for this analysis, and the results are presented in Figure 23. From both SHAP and LIME, the probability of customer churn was identified as the most influential feature for the selected instances. Both models indicated that an increase in the probability of churn had a positive impact on the predicted result. Additionally, there are other similar features listed as the top eight most influential feature according to both model, such as: “hvc_tier_gold”, “avg_3m_pct_change” and “poin”. However, it is important to note that some features showed different impact directions between the two models. For example, the “hvc_tier_gold” feature had a positive impact according to SHAP, but a negative impact according to LIME.



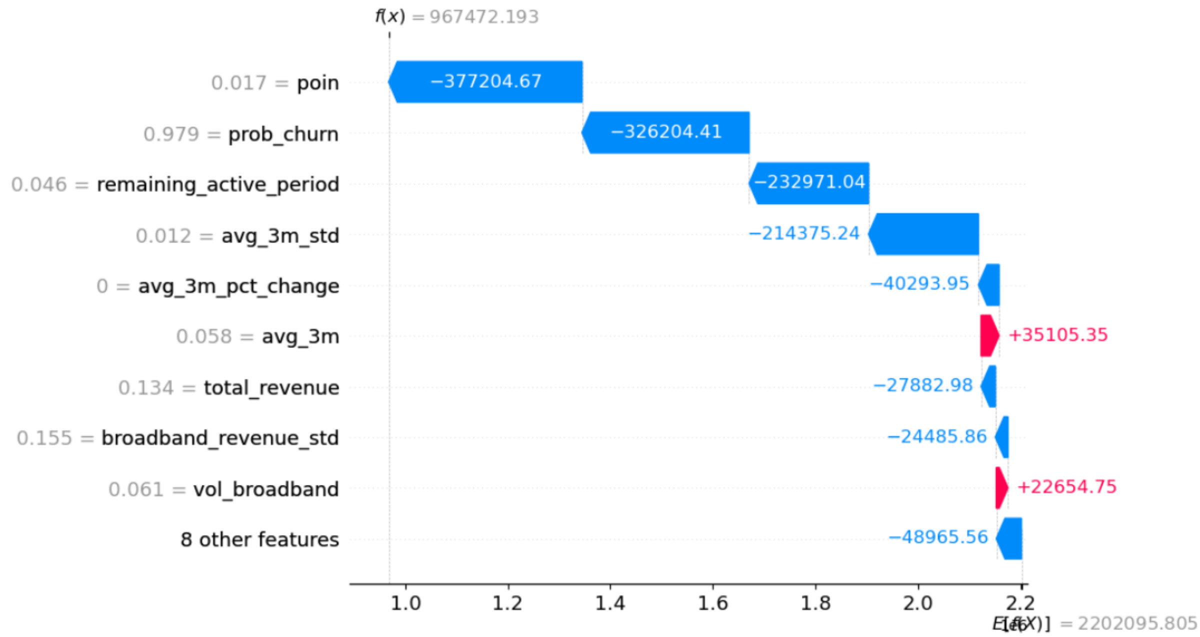
(a)



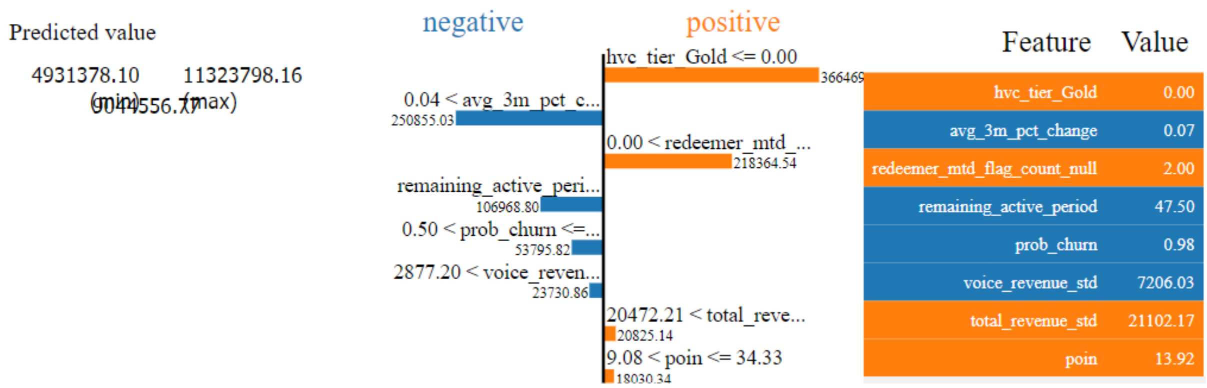
(b)

Figure 23 Approach A.1: (a) SHAP value and (b) LIME value for data sample X

To provide more comprehensive perspective, another data sample, sample Y, was selected for comparison. As sample Y represents instances with a high probability of churn, the analysis aimed to uncover distinct patterns in feature importance that presented in Figure 24. In contrast to sample X, the top five features based on SHAP values showed negative contributions to the predicted value. Specifically, the most important feature, "poin," had a negative contribution of 377,204, followed by "prob_churn" with a high negative contribution of 326,204 to the predicted result. This different could be resulted due to the different value of each feature and the interaction between them. According to LIME, the most important feature for sample Y was "hvc_tier_Gold," which made a positive contribution of 366,469, followed by "avg_3m_pct_change" with a negative contribution of 250,855 to the predicted RAR value.



(a)



(b)

Figure 24 Approach A.1: (a) SHAP value and (b) LIME value for data sample Y

5.3.2.3 Dependence Plot

Dependence plot analysis using SHAP also conducted to find the interaction between the features and SHAP values. For this analysis, “prob_churn” was selected as the main feature and the other top five features, such as 'poin', 'avg_3m_std', 'avg_3m' and 'total_revenue', was selected as the secondary feature. Figure 25 displays the dependence plots, revealing notable insights into the relationship between the probability of customers’ churn and SHAP value. It becomes evident that there is a negative correlation between churn probability and the SHAP value. As the probability of churn increases, the corresponding SHAP value contribution to predicted result is decreases. This finding aligns with the previously discussed correlation between churn probability and RAR in Subsection 5.2.1.

Furthermore, the interaction between the "prob_churn" value and other features revealed distinct impacts on the SHAP values. It was observed that a range of "prob_churn" values could result in the same SHAP value contribution. To delve deeper into this phenomenon, the interaction with another feature was analyzed. From the interaction with “poin” feature, it could be seen that a low probability of churn with high “poin” value tend to have higher SHAP value contribution compared to low probability of churn with low “poin”. However, when the churn probability is high and the “poin” value is high, the SHAP value contribution decreases. This pattern is also found across the dependence plots for the other selected secondary features, namely "avg_3m_std," "avg_3m," and "total_revenue". It is understandable as the “poin” feature represents the historical spending of the customer, and the other features could be considered as the other representation of customer spending. Therefore, all the selected secondary features exhibit similar interaction patterns with the probability of churn.

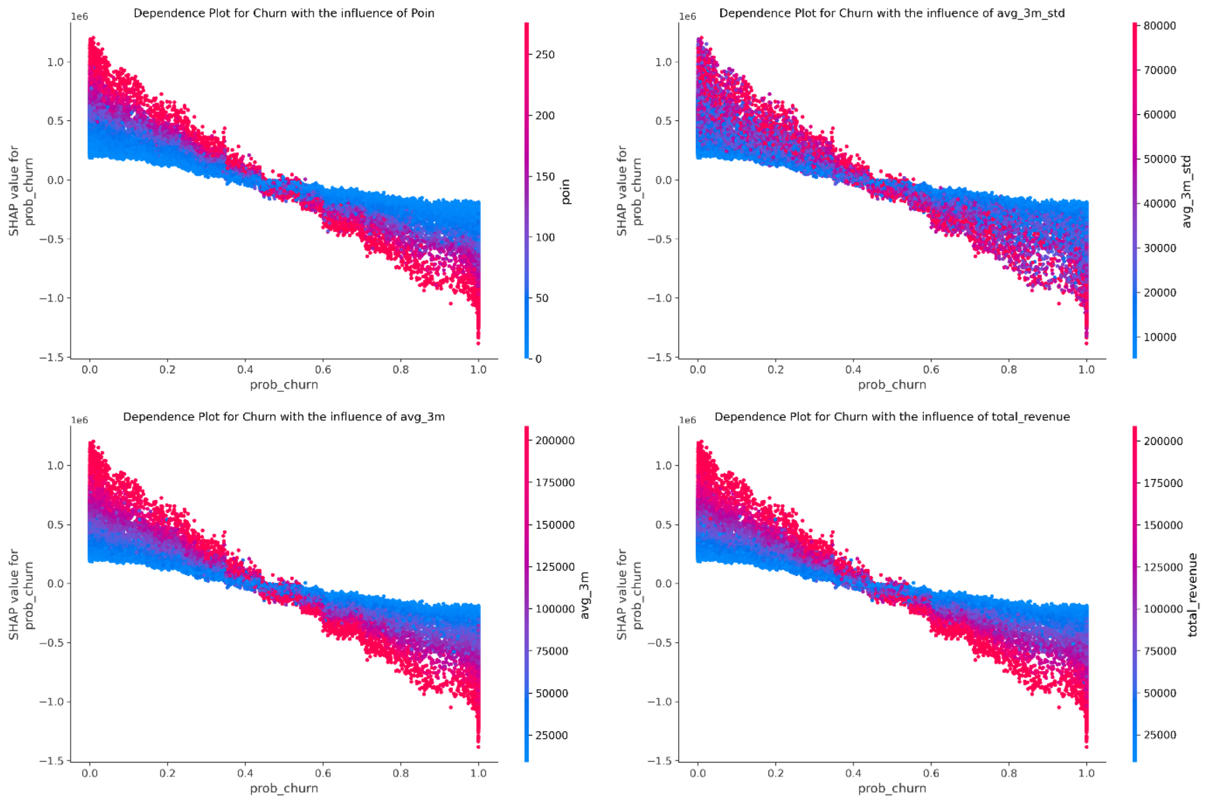


Figure 25 Approach A.1: SHAP Dependence plot of Probability of Churn

5.3.3 Model Validation

To assess the robustness of the final model, various splitting strategies were employed, and the results are presented in Table 11. The table demonstrates that slight variations in performance metrics occur as the split ratio changes from 50:50 to 90:10, but overall, the model's performance remains consistent. Notably, all metrics exhibit similar performance with only negligible differences. The best MAPE and MAE were observed in the 80:20 ratio, while the best RMSE was obtained in the 70:30 ratio. Furthermore, the R^2 values consistently remained high at 0.91 across all splitting strategies. These findings indicate that the model exhibits robustness and stability across different split ratios, demonstrating its ability to generalize well to unseen data.

Table 11 Approach A.1: Final Model Performances Across Different Split strategy

Split Ratio	MAPE	MAE	RMSE	R^2
50:50	19.62	356,987	561,280	0.91
60:40	19.5	355,546	560,821	0.91
70:30	19.32	353,537	559,225	0.91
80:20	19.27	352,344	561,539	0.91
90:10	19.36	354,396	562,349	0.91

Furthermore, the final model also evaluated using 5-fold CV to give detailed understanding of the model validation process and further ensure that the model is not overfitted and has a stable performance. The results are presented at the Table 12 below. As can be observed, the performance on all evaluation metrics is reasonably consistent across the five folds for both the training and test sets. This indicates that the final model is not overfitting as it performs similarly well on unseen data as on the training data. Additionally, the stability of the model is evident, as the performance metrics do not fluctuate significantly across the different folds. This suggests that the model's performance is not heavily reliant on any particular subset of the data, demonstrating its generalizability and robustness.

Table 12 Approach A.1: Final Model 5-fold Cross Validation Result

Fold	Train				Test			
	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE	R^2
1	15.87	290,105	474,393	0.93	19.38	352,691	560,862	0.91
2	15.94	290,944	476,666	0.93	19.28	352,339	552,498	0.91
3	15.74	289,685	476,449	0.93	19.30	354,723	556,717	0.91
4	15.88	288,910	472,150	0.94	19.39	357,390	567,578	0.91
5	15.88	290,527	475,742	0.93	19.54	351,761	553,919	0.91
Avg	**15.86**	**290,034**	**475,080**	**0.93**	**19.38**	**353,781**	**558,315**	**0.91**

The group split method was also employed to examine the model's performance under different splitting strategies, with results presented in Table 13. In line with expectations from traditional cross-validation strategies, the model's performance metrics on the training set were better than those on the test set. While there were slight fluctuations in the test set results, these were not significant. The best MAPE on the test set was achieved by test_group 4, with a rate of 18.94%. The worst MAPE was seen in test_group 3, with a rate of 21.38%. These fluctuations may be attributed to the varied characteristics of data across regions. For instance, region tier 3, which corresponds to the middle-low region, could possess distinct patterns that are not as easily captured by the model trained primarily on data from other regions.

Table 13 Approach A.1: Final Model Group-Split Result

Test	Train				Test			
group	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE	R^2
1	15.84	284,610	467,069	0.94	19.76	402,326	622,598	0.9
2	15.48	289,085	475,847	0.94	19.09	338,876	536,929	0.91
3	15.28	289,954	474,848	0.94	21.38	341,684	541,700	0.91
4	16.17	286,981	467,675	0.93	18.94	518,462	755,218	0.88
Avg	**15.69**	**287,658**	**471,360**	**0.94**	**19.79**	**400,337**	**614,111**	**0.90**

To further evaluate the model's learning efficiency and robustness, the learning curves to evaluate the model performances according to the number of training data used. The metric used in the learning curve is RMSE and the result is presented at Figure 26 below. It could be seen that the training scores increase with the size of the training set, while the Validation_scores decrease. This trend reflects the model's decreasing ability to fit the training data perfectly (as indicated by the increasing training error), but also its improved generalization to unseen data (as shown by the decreasing validation error). Furthermore, the gap between the training and validation scores is relatively small, especially at larger training sizes. This reinforce the notion that the model isn't overfitting as it's generalizing well to unseen data.

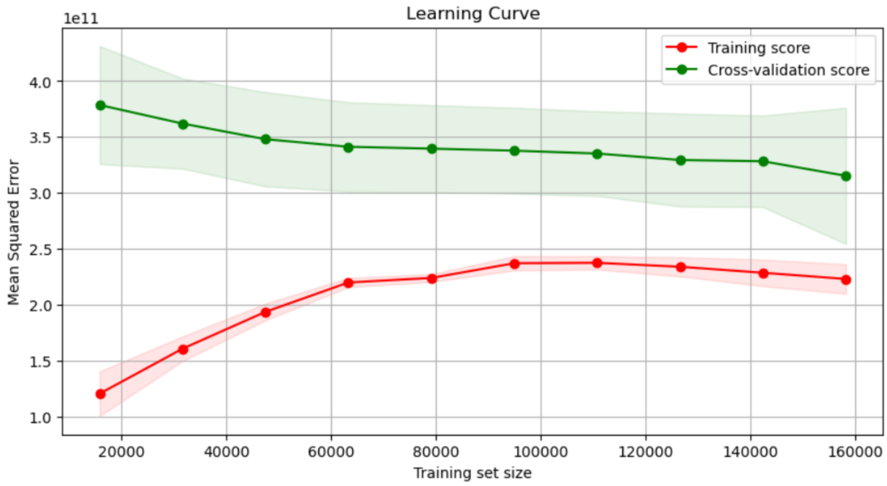


Figure 26 Approach A.1: Final Model Learning Curve Analysis

To validate the model's outputs, a comparative analysis was conducted between the impactful features identified by the ML model and the variables emphasized in traditional calculation methods. This comparative analysis enhances the assessment of the model's coherence and reliability in recognizing essential features. As per Equation 4.1, the discount rate used in RAR calculation for Approach A is derived from the WACC value multiplied by the customer churn probability and divided by the market's expected churn probability. Thus, the customer churn probability is a crucial feature in RAR calculation using the traditional method. Consistently, the final model also identifies the churn probability as an important feature, where it ranked second in importance according to both feature importance and SHAP global explanations. Furthermore, the local explanation from two randomly selected data samples also identified churn probability as the most significant feature as per both SHAP and LIME. Figure 24's dependence plot more clearly demonstrates the relationship between churn probability and SHAP value, revealing a negative correlation. As churn probability increases, the corresponding SHAP value contribution to the predicted RAR decreases and can even turn negative. This finding aligns with the traditional RAR calculation method, as a higher churn probability leads to a higher discount rate, resulting in a lower calculated RAR value.

Additionally, the final model other most important features, “point”, “avg_3m_std”, “avg_3m”, “total_revenue” and “total_revenue_std”, strengthens its coherence between the model with traditional calculation methods. Those features represent various aspects of customer spending or the revenue derived from the customer. This aligns perfectly with the formula used to calculate RAR (Equation 3.1), where the RAR is determined by the sum of the customer's revenue throughout the observation period, divided by the discount rate raised to the power of the period. Moreover, two of these features, “avg_3m_std” and “total_revenue_std”, represent the volatility of customer spending, a factor that is bound to occur given the six semesters of observations. Such volatility inevitably affects the calculated RAR revenue, and the final model captures this by ranking these features among the most important.

5.4 RAR Prediction Model: Approach A.2 (Median WACC value)

The graph presented in Figure 27 portrays the association between the number of features and their respective performances in the feature selection process for Approach A.2. Similar with the result from previous approach, the average MAPE displays fluctuations when the number of features is increased from 5 to 9 and consistently improve afterwards. The MAPE reached 23.85% when the total features are 18 and then the improvement slow smaller change. In contrast, the average RMSE demonstrates a continuous improvement as the number of features increases. This suggests that the incorporation of additional features contributes to a better fit of the model to the observed data, resulting in reduced prediction errors. The average RMSE consistently decreases from 587,396 for 5 features to 545,709 for 18 features.

The selected number of features for the final model in this approach was 18 features. Although the larger number of features could yield improved model performances, a close examination of the graph reveals that the marginal improvement for the number of features beyond 18 are minimal. Hence, based on the observed trend, it can be inferred that 18 features represent the most optimal number for this particular model.

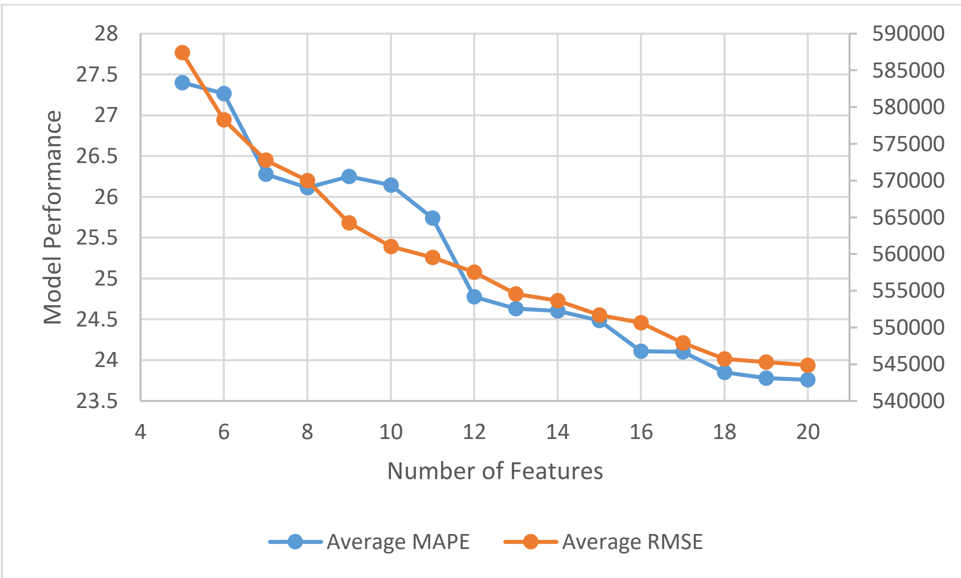


Figure 27 Approach A.2: Number of feature vs model performance (5-fold CV)

5.4.1 *Model Performance Evaluation Metric*

The Table 14 presents the performance metrics of the models at different stages of model development. The Catboost algorithm consistently demonstrates strong performance across all metrics. In the initial model using all features, Catboost achieves a MAPE of 23.21%, an MAE of 349,567, an RMSE of 535,188, and an R^2 of 0.91. Through feature selection, the model's performance remains stable, with slight variations in the performance. Furthermore, hyperparameter tuning using random search leads to improved performance, resulting in a lower MAPE of 20.33%, an MAE of 353,120, an RMSE of 556,118, and a slightly lower R^2 of 0.90. The final model, after grid search hyperparameter tuning, exhibits the best MAPE of 19.61%, the lowest MAE of 346,318, an RMSE of 547,301, and an R^2 of 0.91.

Table 14 Approach A.2: Model Development (5-fold CV)

Model & Model Development Stage	MAPE	MAE	RMSE	R^2
Catboost				
Initial model and all features	23.21	349,567	535,188	0.91
Feature selection	23.85	355,790	545,710	0.91
Random search hyperparameter tuning	20.33	353,120	556,118	0.90
Grid search hyperparameter tuning	**19.61**	**346,318**	**547,301**	**0.91**
XGBoost				
Initial model and all features	24.21	364,263	559,590	0.904
Feature selection	23.90	365,209	561,521	0.904
Random search hyperparameter tuning	22.91	359,090	551,957	0.902
Grid search hyperparameter tuning	22.21	353,787	548,270	0.91
Random Forest				
Initial model and all features	24.52	373,964	577,500	0.9
Feature selection	24.47	375,088	577,966	0.9
Random search hyperparameter tuning	24.44	374,299	576,814	0.9
Grid search hyperparameter tuning	23.59	372,164	574,789	0.9

In comparison, the XGBoost algorithm also demonstrates competitive performance. Similar to Catboost, XGBoost achieves consistent performance across the different stages of model development. The initial model using all features yields a MAPE of 24.21%, an MAE of 364,263, an RMSE of 559,590, and an R^2 of 0.904. Through feature selection and hyperparameter tuning, the model's performance remains relatively stable, with slight improvements observed in the MAPE (23.90% and 22.91%, respectively), MAE, RMSE, and R^2.

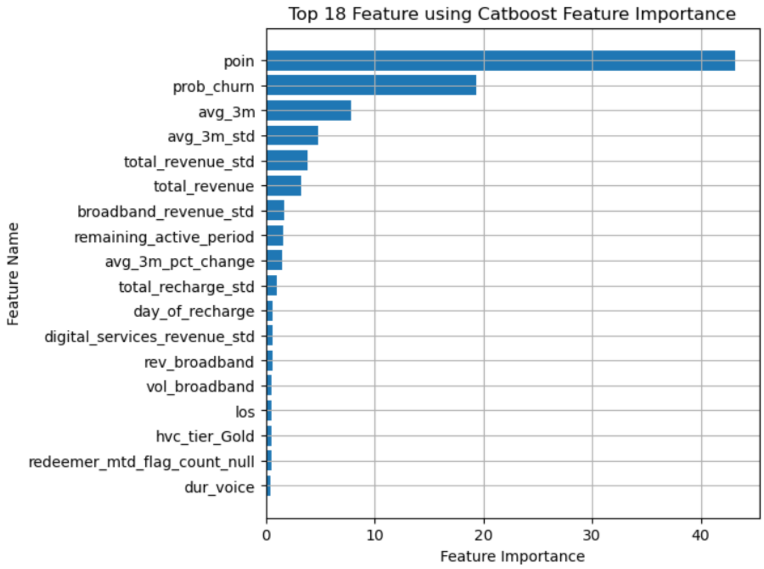
In contrast, the Random Forest algorithm demonstrates comparatively lower performance across all model development stages. The initial model using all features yields a MAPE of 24.52%, an MAE of 373,964, an RMSE of 577,500, and an R^2 of 0.9. Despite feature selection and hyperparameter tuning, the model's performance shows minimal improvement, resulting in a final model with a MAPE of 23.59%, an MAE of 372,164, an RMSE of 574,789, and an R^2 of 0.9.

Overall, Catboost consistently demonstrates the best performance across all model development stages, with the lowest MAPE, RMSE, and highest R^2 values. XGBoost and Random Forest exhibit comparable performance but fall slightly behind Catboost in terms of the overall model performance. This resulted in Catboost also selected as the final model for this particular approach.

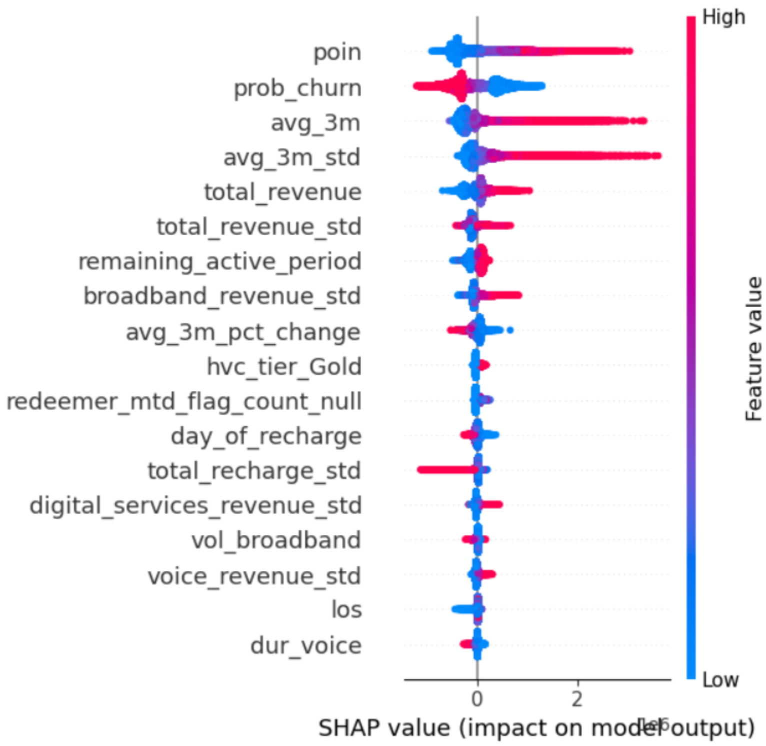
5.4.2 Model Interpretation and Feature Importance Analysis

5.4.2.1 Global Explanation

Feature importance analysis was conducted on the best model using two methods: the embedded feature importance method within the model and the SHAP explanation using SHAP smmary plot. The results that presented in Figure 28 revealed consistent findings with the previous approach. The most important feature, "poin", exhibited a feature importance value more than double that of the second most important feature, "prob_churn". Similarly, the third most important feature, "avg_3m", had less than half the feature importance value of "prob_churn". The evaluation using SHAP values corroborated the findings from the feature importance method. The list of most important features remained similar, with "poin" retaining its position as the most influential feature. However, the difference in SHAP value impact between the most important feature and the other top four features is not as significant as the difference observed in feature importance.



(a)

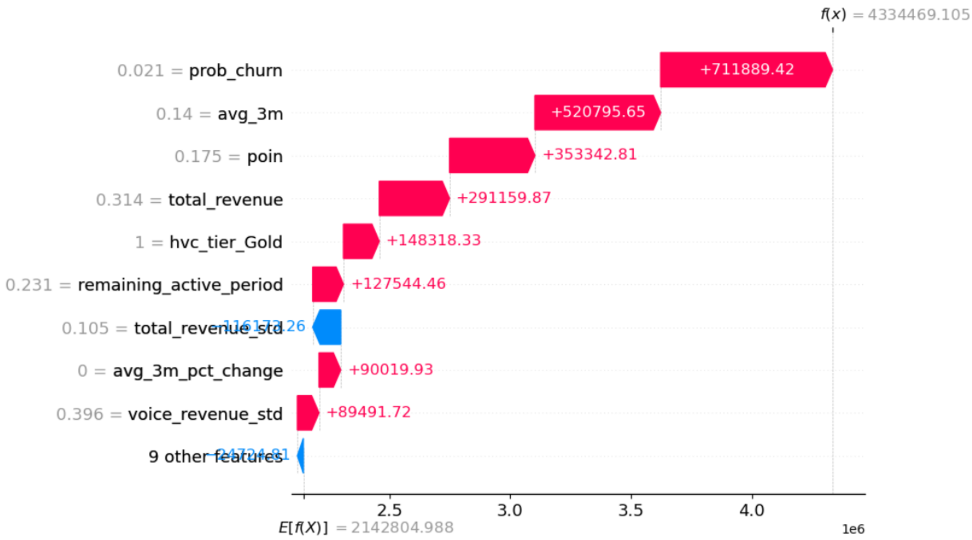


(b)

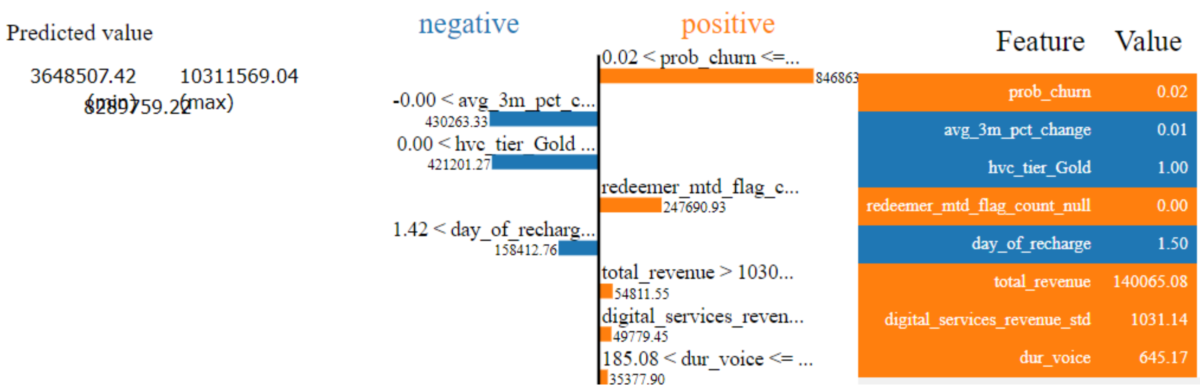
Figure 28 Approach A.2: (a) Feature Importance and (b) SHAP value

5.4.2.2 Local Explanation

To gain a more detailed understanding of the impact of each feature, a local explanation was performed using both SHAP value and LIME. The same data sample utilized in the previous approach was selected for this analysis. The results from both SHAP value and LIME, which presented in Figure 29, consistently demonstrated that the top four features exhibited a similar impact direction to the previous approach, although the magnitude of the impact varied. Notably, the most important feature according to both SHAP value and LIME was "prob_churn," which had a positive impact on the predicted result. However, the remaining features display varying rankings and even different impact directions between SHAP value and LIME. For instance, the second most important feature according to SHAP value, "avg_3m", is not among the top eight most important features according to LIME. In contrast, LIME identifies "avg_3m_pct_change" as the second most important feature, which is closely related to "avg_3m", while giving different impact direction with SHAP value. Moreover, despite "poin" being identified as the most important feature according to SHAP global explanation, it ranks lower in importance according to SHAP value and does not even appear among the top eight important features according to LIME for this specific instance.



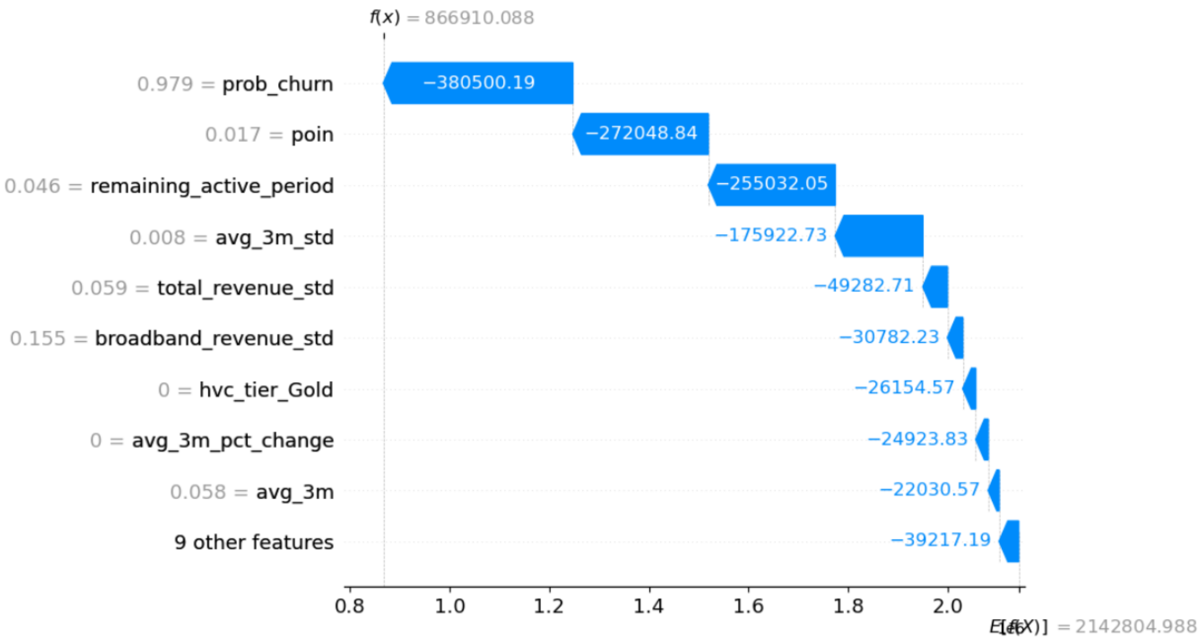
(a)



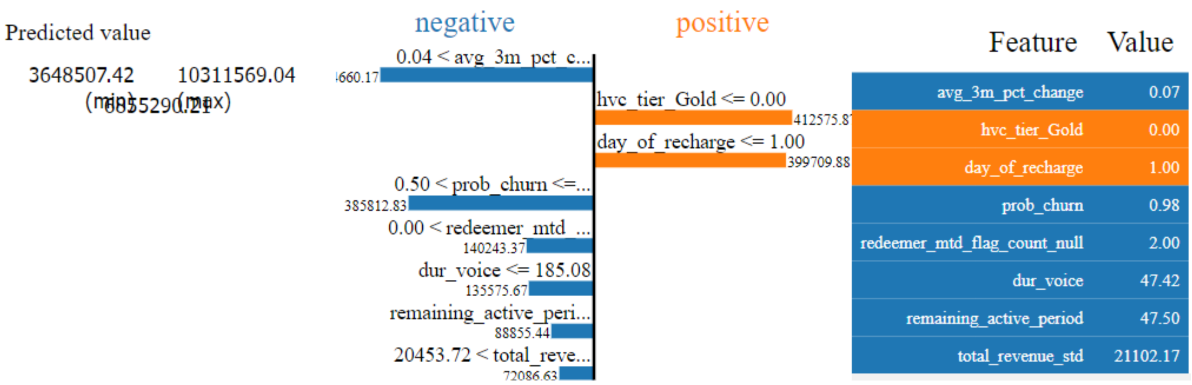
(b)

Figure 29 Approach A.2: (a) SHAP value and (b) LIME value for data sample X

Furthermore, data sample Y was used as the second sample for comparison. Interestingly, all the top features had negative contributions to the predicted SHAP values. According to SHAP, the most important feature was "prob_churn" which negatively contributing 380,500, followed by "poin" with a negative contribution of 272,048. Conversely, according to LIME, the most important feature was "avg_3m_pct_change" with a negative contribution of 444,660, while "hvc_tier_Gold" made a positive contribution of 412,575. Another noteworthy feature was "day_of_recharge," ranking third in importance and making a relatively high positive contribution of 399,709. These contrasting results highlight the differences in how SHAP values and LIME interpret and attribute importance to the features for individual instances. The detailed result is presented in Figure 30 below.



(a)



(b)

Figure 30 Approach A.2: (a) SHAP value and (b) LIME value for data sample Y

5.4.2.3 Dependence Plot

The dependence plot of “prob_churn” with the other top five features (“poin”, “avg_3m”, and “avg_3m_std”) was also examined and presented in Figure 31. The findings mirrored those of the previous approach, with a negative correlation observed between the probability of churn and the corresponding SHAP value. The interaction between "prob_churn" and the other features followed a similar pattern, whereby instances with a low probability of churn and a high "poin" value tended to exhibit a higher SHAP value contribution compared to those with a low "poin" value. However, when both the churn probability and the "poin" value were high, the SHAP value contribution decreased.

Interestingly, the impact of "avg_3m_std" on the SHAP value, in relation to the probability of churn, was not as straightforward as the other features. There were instances where a high "avg_3m_std" value combined with a low probability of churn still resulted in a relatively low SHAP value. This phenomenon can be attributed to the nature of "avg_3m_std" that slightly different with the other features. While the other features represent the historical customer spending, for example the "avg_3m" measures the average revenue from customer in the last 3 months, the "avg_3m_std" measures the standard deviation of such value. Therefore, it introduces additional complexities to its interaction with the “prob_churn”.

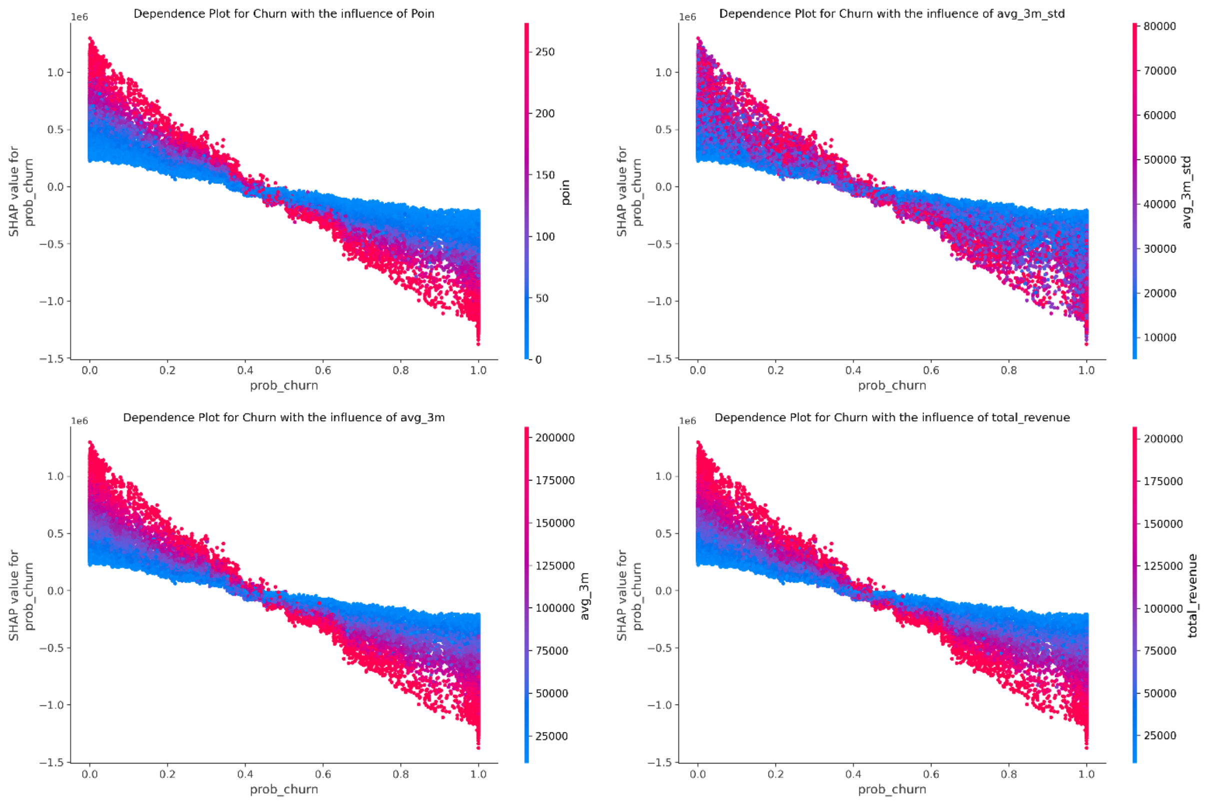


Figure 31 Approach A.2: SHAP Dependence plot of Probability of Churn

5.4.3 Model Validation

The best model from approach A.2 then validated using different splitting strategies to assess its performance. The splitting ratio ranges from 50:50 to 90:10 and the results presented in the Table 15. The result indicates the stable performance across splitting ratio, where the best performance is achieved by 90:10 splitting ratio. These findings indicate that the model exhibits robustness and stability across different split ratios, demonstrating its ability to generalize well to unseen data.

Table 15 Approach A.2: Final Model Performances across Different Split Strategies

Split Ratio	MAPE	MAE	RMSE	R^2
50:50	20.14	352,299	559,715	0.91
60:40	19.75	348,603	554,279	0.91
70:30	19.7	349,510	556,280	0.91
80:20	19.86	347,267	552,860	0.91
90:10	19.64	344,555	546,625	0.91

Table 16, showing the result of a 5-fold CV process, serves as further evidence to the validity of the developed model. Similar with the previous approach, performance consistency across all evaluation metrics and folds for both the training and test sets becomes apparent. This consistency indicates effective handling of overfitting and equal performance with unseen and training data by the model. Stability of the model is also visible from the minor fluctuation of performance metrics across different folds, suggesting minimal influence of any specific subset of the data on the model's performance.

Table 16 Approach A.2: Final Model 5-fold CV Results

Fold	Train				Test			
	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE	R^2
1	16.3	287,640	469,430	0.93	19.77	347,029	551,947	0.91
2	16.25	289,099	472,341	0.93	19.64	346,259	550,746	0.91
3	16.36	286,834	469,160	0.93	19.5	343,313	540,605	0.91
4	16.32	286,427	468,447	0.93	19.41	346,215	543,951	0.91
5	16.33	287,466	469,619	0.93	19.73	348,773	549,256	0.91
Avg	**16.31**	**287,493**	**469,799**	**0.93**	**19.61**	**346,318**	**547,301**	**0.91**

The group split analysis, as shown in Table 17, yielded results similar to the previous approach, with slight variations in performance across the splits. Notably, test_group 4 achieved the best MAPE, but also exhibited the highest RMSE and lowest R^2 value. Conversely, test_group 3 consistently had the worst MAPE, consistent with the previous approach. It is worth noting that the relatively poorer performance in MAE and RMSE for test_group 4 may be caused by its inclusion of regions with lower economic power, potentially limiting the model's ability to generalize effectively to these specific regions. However, despite these slight variations in error rates, the model's overall performance remains robust across all regions.

Table 17 Approach A.2: Final Model Group-Split Results

Test group	Train				Test			
	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE	R^2
1	16.29	282,609	464,647	0.93	20.1	393,836	607,354	0.9
2	15.98	286,177	469,020	0.94	19.34	332,385	526,891	0.91
3	15.7	287,718	471,364	0.93	21.62	334,331	530,774	0.91
4	16.57	284,549	463,970	0.93	19.02	508,967	744,998	0.87
Avg	**16.14**	**285,263**	**467,250**	**0.93**	**20.02**	**392,380**	**602,504**	**0.90**

Another critical aspect of model validation is the evaluation of learning efficiency and robustness using learning curves. Figure 32 presents the learning curves using RMSE as the evaluation metric. Similar pattern with the approach A.1 also appears as with the expansion of the training set size – an increase in the training scores and a decrease in the validation scores are occurs. This trend demonstrates the model's aptitude in generalizing unseen data, as it becomes less efficient at fitting the training data perfectly. The relatively narrow gap between the training and validation scores, especially at larger training sizes, strengthens the idea that the model is not overfitting.

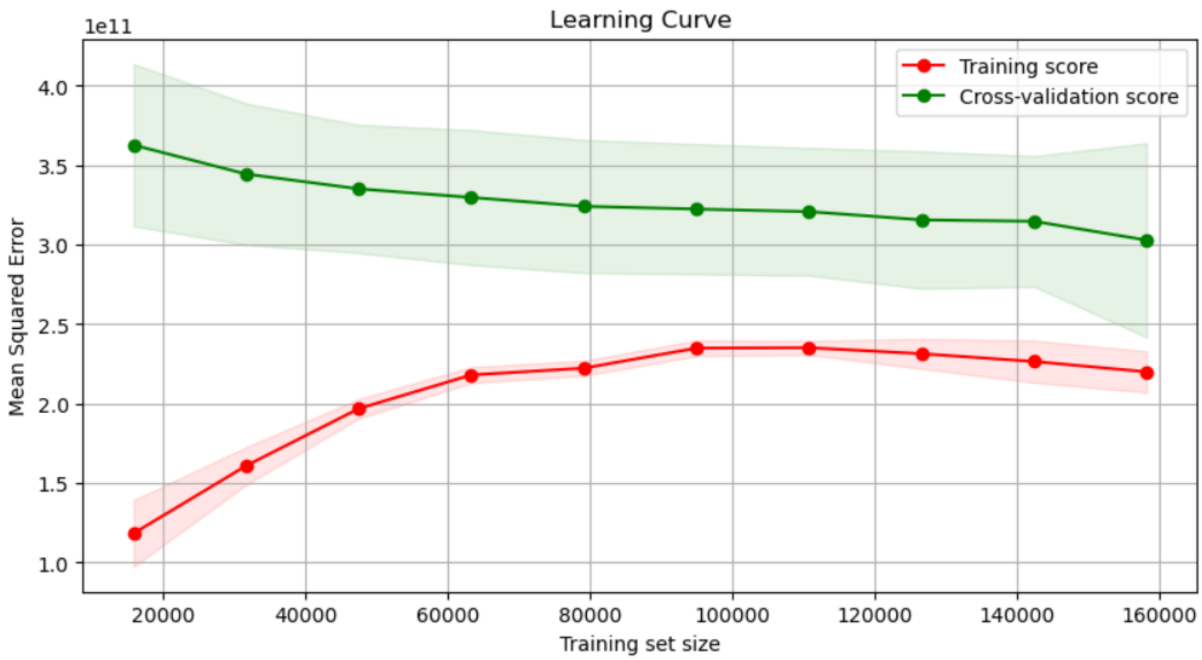


Figure 32 Approach A.2: Final Model Learning Curve Analysis

Model outputs also undergo validation through a comparative study between influential features as determined by the ML model and variables emphasized in traditional calculation methods. Just like Approach A.1, customer churn probability remains an influential feature in the traditional calculation of RAR, and the ML model maintains its importance. It is evidence that in both the global explanation and the local explanation, the probability of churn remains as the top important feature. On the dependence plot (Figure 30), the significance of SHAP value contribution according to the probability of churn also have negative correlation, where the higher probability of churn, the lower its SHAP value.

Furthermore, the model for Approach A.2 also exhibits striking similarities with the conventional calculation methods, as evidenced by the importance assigned to its top features. Like the previous approach, "poin", "avg_3m_std", "avg_3m", "total_revenue", and “total_revenue_std” again emerge as critical. These features, in essence, reflect the varying dimensions of customer spending and their revenue. This interpretation conforms with the RAR calculation formula (Equation 3.1), adding weight to the model's credibility. Intriguingly, features such as “avg_3m_std” and “total_revenue_std” stand for the fluctuations in customer spending over the six semesters of observation. These movement inevitably influence the RAR value, and these feature captured as the most important feature from the model.

5.5 RAR Prediction Model: Approach A.3 (Maximum WACC value)

Figure 33 presents the results of evaluating the model's performance in relation to the number of features used using MAPE and RMSE as the evaluation metric. Similar with the previous approach, the MAPE showed performance fluctuations throughout the observation, while the RMSE showed constant improvement. The optimal number of features for the final model was determined to be 17. This choice was based on the trade-off between model performance and complexity. It was observed that increasing the number of features beyond 17 did not result in significant improvement in the average MAPE and average RMSE. Therefore, selecting 17 features was deemed the most effective approach, striking a balance between model accuracy and complexity.

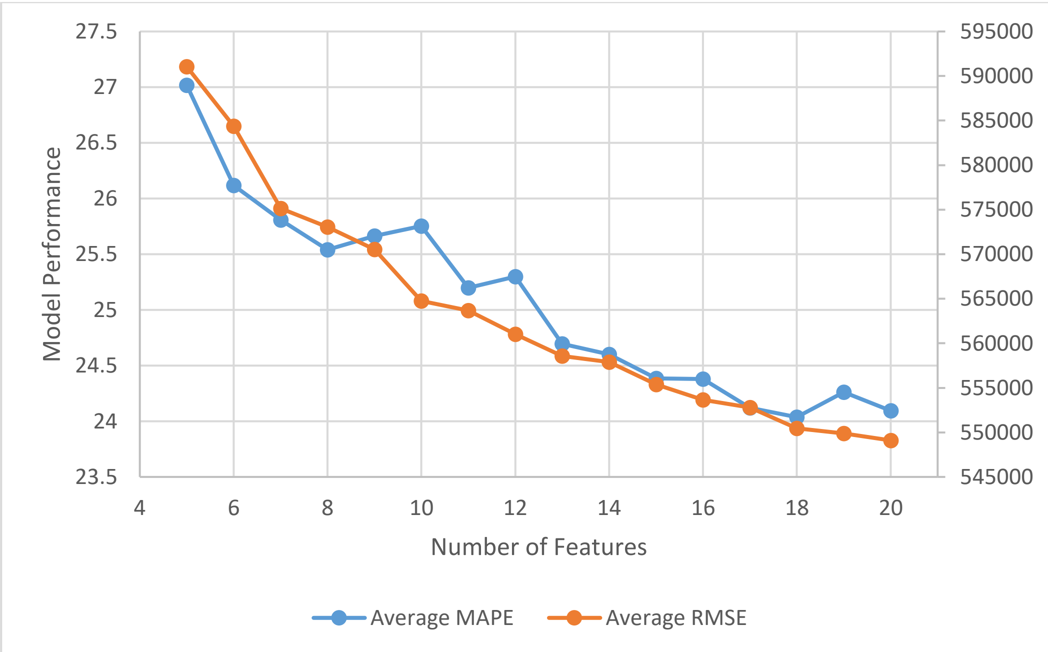


Figure 33 Approach A.3: Number of feature vs model performance (5-fold CV)

5.5.1 Model Performance Evaluation Metric

Table 18 presents the performance results of the models across various development stages for Approach A.3. The Catboost model, initially trained with all features, achieved a MAPE of 23.86%, an MAE of 352,323, an RMSE of 540,175, and an R^2 value of 0.91. A minor decline in the MAPE to 23.35% was observed following the feature selection process, while a modest uptick in error was noted in other metrics. The employment of hyperparameter tuning technique further enhanced the model's performance, decreasing the MAPE to 19.92% in random search and the subsequent grid search led to an even lower MAPE of 19.80%. In contrast, the MAE and RMSE displayed mixed trends - increasing with random search tuning, then decreasing with grid search while the R^2 remained relatively steady at 0.91 during the process.

Table 18 Approach A.3: Model Development

Model & Model Development Stage	**MAPE**	**MAE**	**RMSE**	R^2
Catboost				
Initial model and all features	23.86	352,323	540,175	0.91
Feature selection	23.35	352,475	541,102	0.91
Random search hyperparameter tuning	19.92	351,129	554,663	0.906
Grid search hyperparameter tuning	**19.80**	**348,060**	**550,109**	**0.908**
XGBoost				
Initial model and all features	24.14	363,165	558,302	0.908
Feature selection	24.03	364,693	559,999	0.902
Random search hyperparameter tuning	22.75	354,497	547,482	0.91
Grid search hyperparameter tuning	22.19	353,865	548,878	0.91
Random Forest				
Initial model and all features	24.19	373,054	576,709	0.9
Feature selection	23.96	374,254	576,936	0.9
Random search hyperparameter tuning	23.55	372,967	576,047	0.9
Grid search hyperparameter tuning	23.39	371,705	574,476	0.9

The subsequent model, XGBoost, achieved a slightly higher MAPE of 24.14% compared to the initial Catboost model, while the remaining metrics also slightly underperformed in relation to Catboost. Following the feature selection process, a slight decrease in the MAPE to 24.03% was recorded, with a minor rise in error across other metrics. The model improved notably with random search hyperparameter tuning, lowering the MAPE to 22.75%, and the grid search further improved this to 22.19%. However, as with the Catboost model, the MAE and RMSE fluctuated, while the R^2 remained stable at 0.91. Despite improvements across all metrics relative to the initial model with all features, the final XGBoost model lagged slightly behind the final Catboost model.

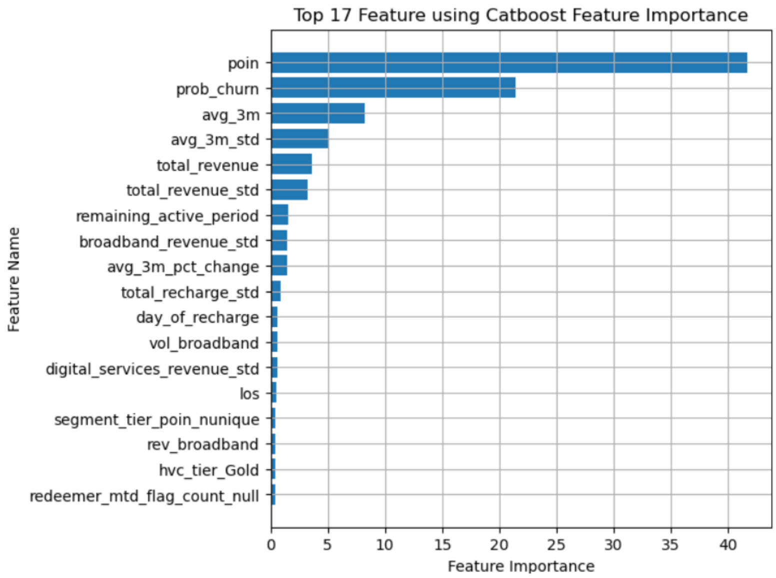
The initial Random Forest model provided competitive MAPE results at 24.19% compared to other models, albeit with higher values for MAE and RMSE. However, improvements across the model development stages were minimal. Hyperparameter tuning achieved slight enhancements in metrics, with the final model registering a MAPE of 23.39%, an MAE of 371,705, and an RMSE of 574,476. While there was an improvement in performance across all metrics relative to the initial model, the final Random Forest model underperformed when compared to the other two models.

In comparing the three models, Catboost and XGBoost consistently surpassed the Random Forest model in all performance metrics. The Catboost model achieved the lowest final MAPE of 19.80% and MAE of 348,060 after grid search hyperparameter tuning. However, XGBoost demonstrated marginally superior RMSE and R^2 values compared to Catboost, albeit with minor differences. Considering the overall performance and the significant improvement in MAPE, the Catboost model was chosen as the final model.

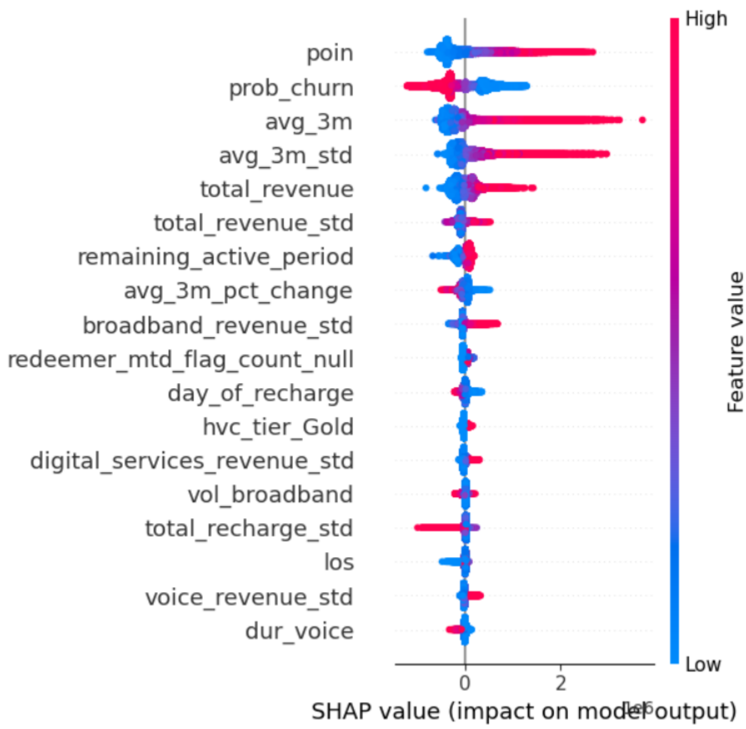
5.5.2 Model Interpretation and Feature Importance Analysis

5.5.2.1 Global Explanation

In terms of feature importance and SHAP global explanation, the top 9 most important features remain consistent with the previous approach. From the Figure 34, could be observed that "poin" emerges as the most influential feature, with a significant margin over the second most important feature, "prob_churn." The SHAP summary plot aligns with this finding, showing that the impact of "poin" on the predicted result is notable but not as distinct as indicated by the feature importance values. "Prob_churn" and "avg_3m" retain their positions as the second and third most important features according to SHAP value. This pattern is similar to the previous approach.



(a)



(b)

Figure 34 Approach A.3: (a) Feature Importance and (b) SHAP value

5.5.2.2 Local Explanation

In the local explanation using both SHAP and LIME in the Figure 35, the most important feature remains consistent with the previous approaches, which is "prob_churn". However, the rankings and contribution directions of the other features differ between the two methods. The “prob_churn” is contributing 738,068 of SHAP value to the predicted result. An interesting finding in this approach is that "redeemer_mtd_flag" emerges as the second most important feature according to LIME, having similar contribution value to "prob_churn", where the later contribute 1,138,101 and the former contribute 1,096,596 to the predicted result. However, according to SHAP value, "redeemer_mtd_flag" does not appear among the top nine most important features.

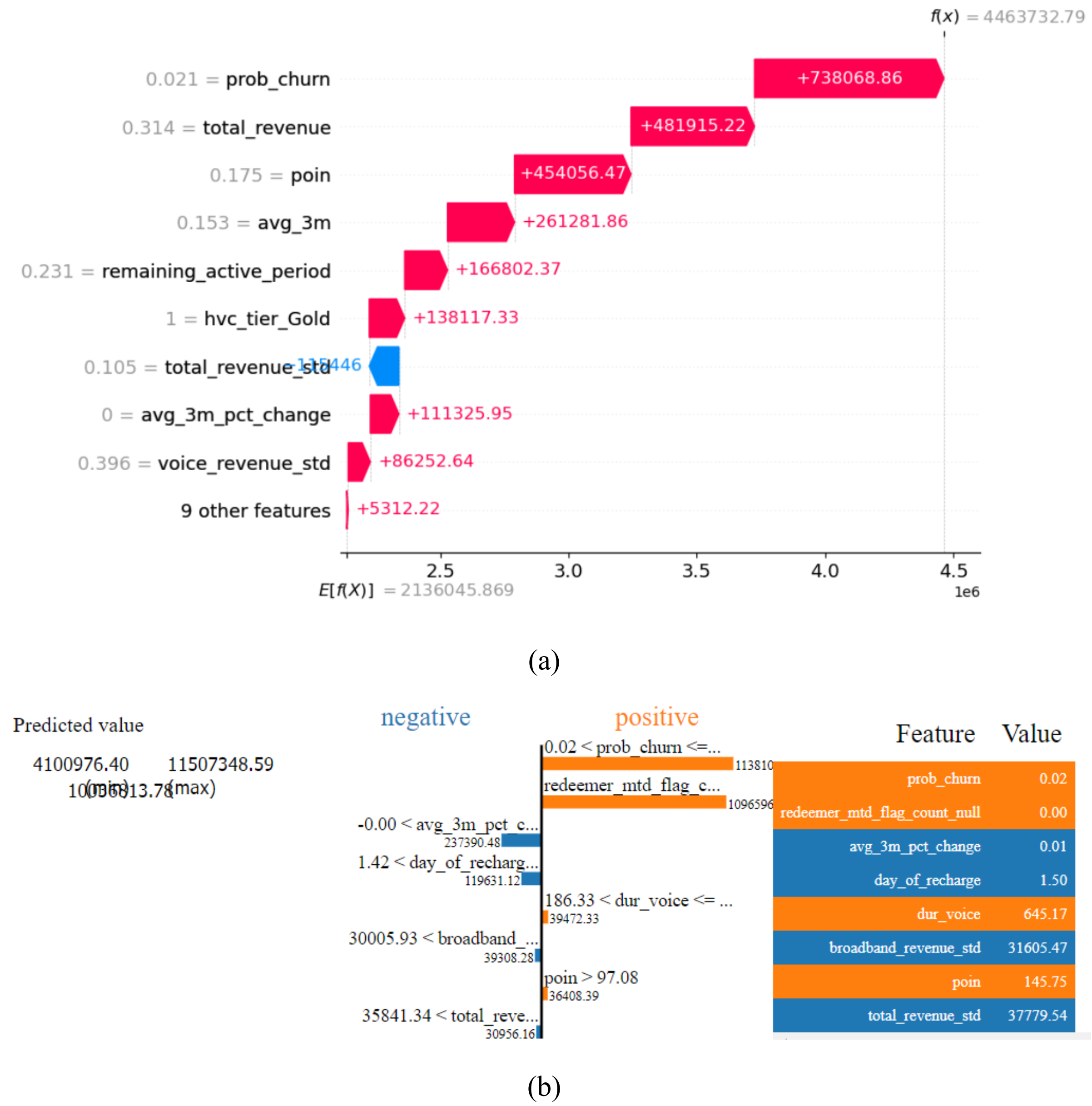
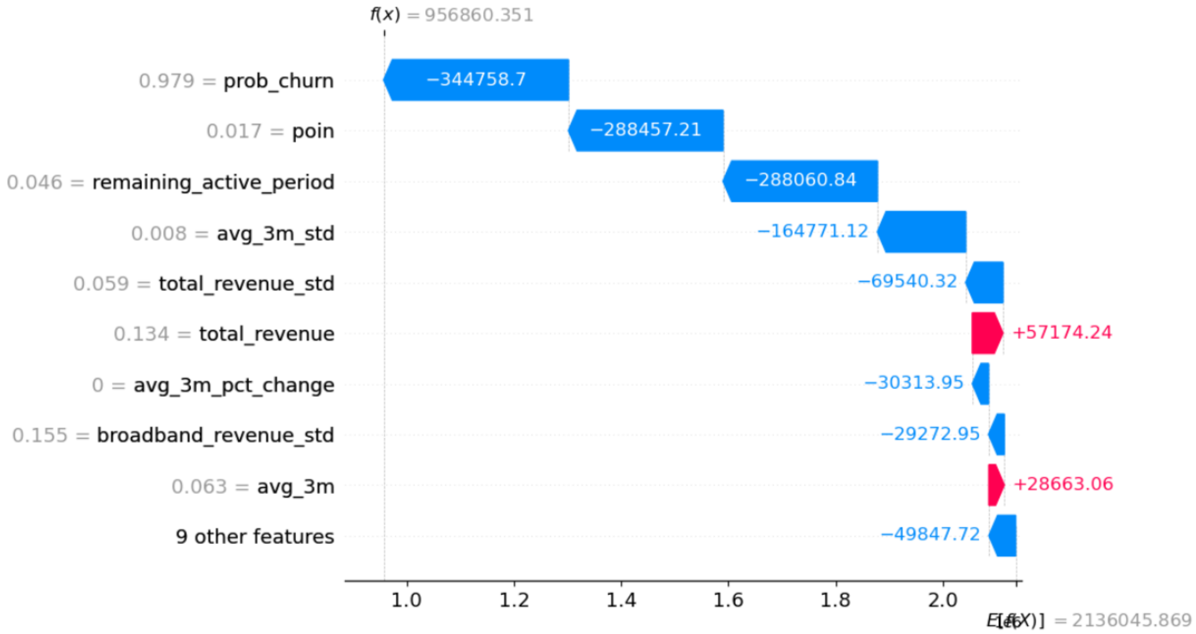
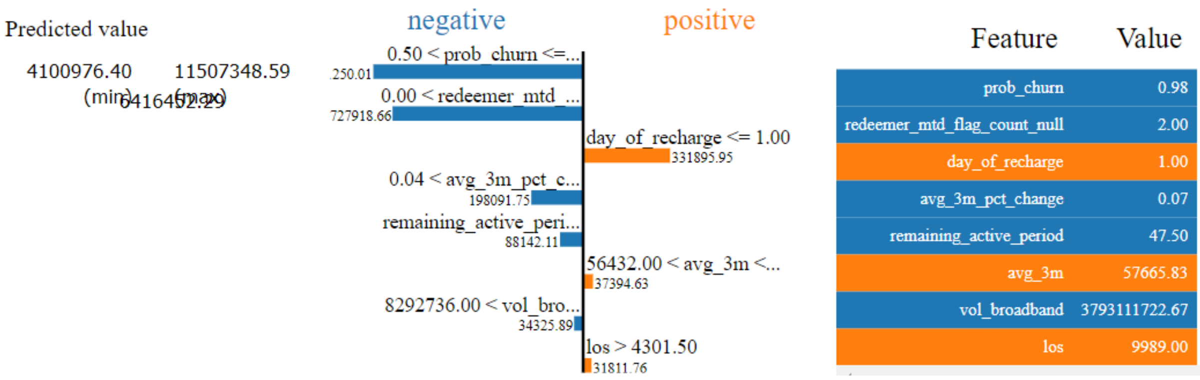


Figure 35 Approach A.3: (a) SHAP value and (b) LIME value for data sample X

Similarly, for sample Y, “prob_churn” was also regarded as the most important feature according to SHAP and LIME (Figure 36). Aside from that, the other significant features according to both of them are different. According to SHAP, the top three most important features are prob_churn, poin and remaining active period which all of them gave negative contribution of 344,758, 288,457 and 288,060 of SHAP value respectively. On the other hand, the top three most important features according to LIME are “prob_churn”, “redeemer_mtd_flag_count_n” and “day_of_recharge” with contribution -801,250, -727,918 and 331,895.



(a)



(b)

Figure 36 Approach A.3: (a) SHAP value and (b) LIME value for data sample Y

5.5.2.3 Dependence Plot

The dependence plot analysis presented in Figure 37 further confirms the consistency of the results across multiple approaches. Similar to the previous findings, the probability of churn exhibited a negative correlation with the SHAP value, demonstrating a consistent pattern of influence on the predictions. The interaction with the other top 5 features also displayed similar trends, indicating that the impact of these features on the SHAP value remains consistent across different approaches. Furthermore, it is important to note that the similarity in results can be attributed to the shared methodology used to calculate the RAR, with variations only in the WACC value. This shared approach ensures that the interaction and impact of each feature on the SHAP value are likely to be similar across the different approaches.

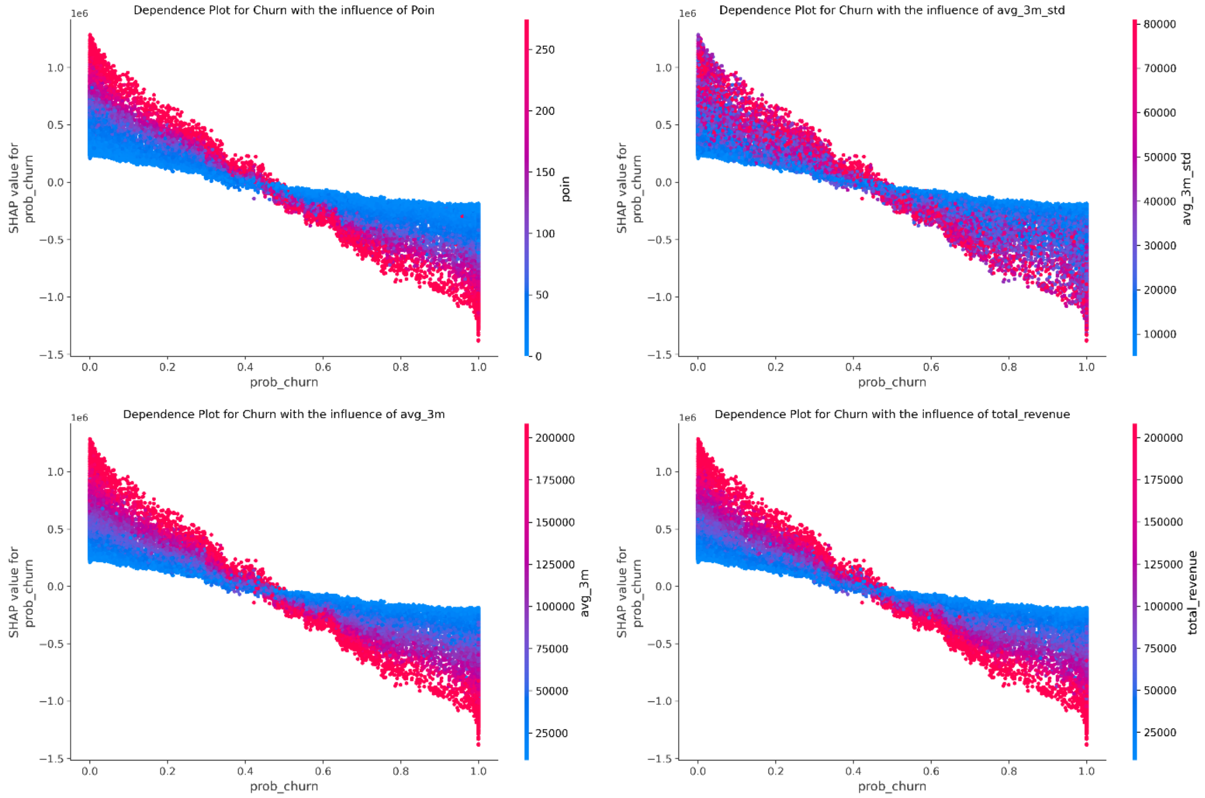


Figure 37 Approach A.3: SHAP Dependence plot of Probability of Churn

5.5.3 Model Validation

The best model from approach A.3 then validated using different splitting strategies to assess its performance. The splitting ratio ranges from 50:50 to 90:10 and the results presented in the Table 19. The result indicates the stable performance across splitting ratio, where the best performance is achieved by 90:10 splitting ratio. These findings indicate that the model exhibits robustness and stability across different split ratios, demonstrating its ability to generalize well to unseen data.

Table 19 Approach A.3: Final Model Across Different Split Strategies

Split Ratio	MAPE	MAE	RMSE	R^2
50:50	19.96	348,378	547,198	0.91
60:40	19.87	346,241	543,214	0.91
70:30	19.58	344,013	538,791	0.91

80:20	19.49	342,713	539,018	0.91
90:10	19.21	341,508	539,892	0.91

The result of the 5-fold CV process for this approach can be seen in Table 20. Similar to the previous models, the consistency of performance on all evaluation metrics across all five folds for both the training and test sets can be noted. This denotes that the model has successfully mitigated overfitting, performing equally well on both the training data and unseen data. The performance metrics' minor fluctuations across the folds further validate the model's stability, signifying that the model's predictions are not significantly influenced by any particular subset of the data.

Table 20 Approach A.3: Final Model 5-fold CV Results

Fold	Train				Test			
	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE	R^2
1	15.41	264,830	438,019	0.94	19.5	343,168	540,756	0.91
2	15.44	266,646	442,098	0.94	19.71	344,521	540,601	0.91
3	15.4	264,617	437,423	0.94	19.72	347,818	550,107	0.91
4	15.34	263,399	436,265	0.94	19.63	352,027	557,146	0.91
5	15.2	265,289	439,618	0.94	19.84	348,501	555,104	0.91
Avg	**15.36**	**264,956**	**438,685**	**0.94**	**19.68**	**347,207**	**548,743**	**0.91**

The evaluation of the model using group split analysis further confirmed its robust performance across different groups. The results are presented in Table 21. Consistent with the previous approaches, there were minor fluctuations in the performance metrics, indicating overall robustness across all groups. Notably, train_group 3 exhibited the highest MAPE value at 22.15%, while train_group 4 had the highest MAE and RMSE values, remaining consistent with previous observations.

Table 21 Approach A.3: Final Model Group Split Results

Test group	Train				Test			
	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE	R^2
1	15.35	260,924	434,420	0.94	20.1	393,687	608,595	0.89
2	14.95	259,692	433,220	0.95	19.43	332,649	527,872	0.91
3	14.71	262,135	436,194	0.94	22.15	334,709	530,085	0.91
4	15.63	264,314	436,086	0.94	19.42	510,876	745,660	0.87
Avg	**15.16**	**261,766**	**434,980**	**0.94**	**20.28**	**392,980**	**603,053**	**0.90**

To evaluate the learning efficiency and robustness of the model under Approach A.3, learning curves are employed. Figure 38 demonstrates these learning curves, utilizing RMSE as the chosen metric. The progression of the training set size elucidates an upward trend in the training scores and a downward trend in the validation scores. This is indicative of the model's improving capacity to generalize unseen data, while it becomes slightly less precise at fitting the training data. The compact gap between the training and validation scores, specifically at larger training sizes, strengthen the idea that the model is not overfitting.

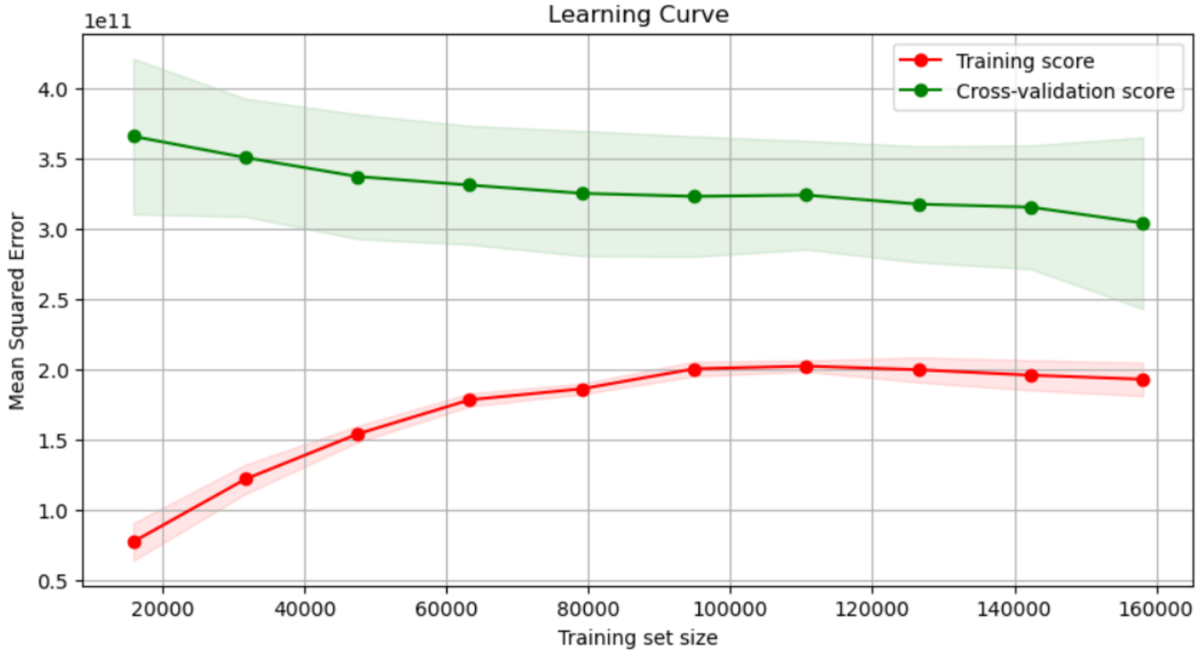


Figure 38 Approach A.3: Final Model Learning Curve Analysis

Validating the model's outputs is also crucial, and a comparative analysis is performed between the significant features recognized by the ML model and the variables emphasized in traditional calculation methods. Much like in the previous approaches, the probability of churn remains a crucial feature in the traditional RAR calculation, and the final model selected in Approach A.3 continues to acknowledge its significance, in both global and local explanation. The relationship between probability of churn with the SHAP value on the dependence plot analysis (Figure 36) also remain similar as they have negative correlation.

A robust link between the ML model and traditional calculation methods strengthen further by ascribing significance to features like “poin”, “avg_3m_std”, “avg_3m”, “total_revenue”, and “total_revenue_std”. Each of these features offers a unique perspective into customer spending patterns and their related revenues, conforming with the RAR calculation formula (Equation 3.1). Furthermore, the model also recognise “avg_3m_std” and "total_revenue_std" as top important features, representing customer spending volatility across the observed six semesters. The model's ability to factor in this volatility while calculating the RAR value mirrors the traditional methods and reinforces its own relevance and reliability.

5.6 RAR Prediction Model: Approach B (Beta value)

Figure 39 provides a comprehensive evaluation of the model's performance in terms of MAPE and RMSE across different numbers of features. The results demonstrate consistent improvements in both metrics as the number of features increases. Starting with 5 features, the MAPE gradually decreases from 25.08% to its lowest point of 21.51% at 17 features. This substantial reduction in MAPE highlights the enhanced predictive accuracy achieved by incorporating a larger set of features. Similarly, the RMSE shows a consistent decline from 709,734 to 659,678 when using 17 features. Beyond 17 features, the improvement in RMSE becomes marginal, indicating diminishing returns in reducing the prediction error. Based on these findings, it is evident that the selection of 17 features represents the optimal configuration for this approach. This choice strikes a balance between maximizing predictive accuracy, as evidenced by the lowest MAPE, and avoiding unnecessary complexity.

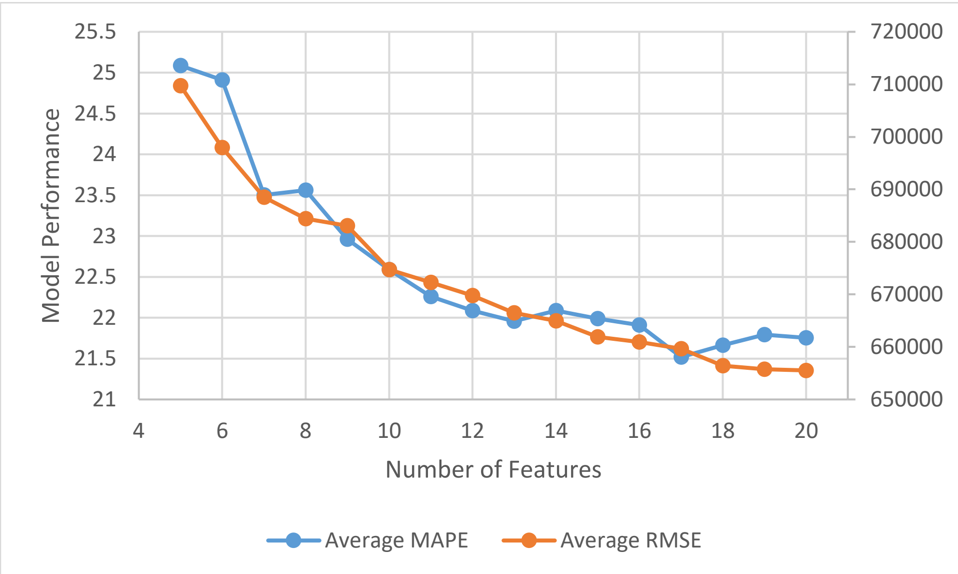


Figure 39 Approach B: Number of feature vs model performance (5-fold CV)

5.6.1 Model Performance Evaluation Metric

Table 22 provides a comprehensive overview of the model performance at different stages of development for Approach B. The Catboost model, initially trained with all features, achieved a MAPE of 21.77%, an MAE of 421,813, an RMSE of 647,108, and an R^2 of 0.92. After feature selection, a slight improvement in MAPE was observed, reducing it to 21.44%.

However, the other metrics showed a minor increase in error. The random search hyperparameter tuning method further reduced the MAPE to 18.67%, and the subsequent grid search yielded an even lower MAPE of 18.35%, indicating improved prediction accuracy. However, the MAE and RMSE exhibited mixed results, increasing with random search but decreasing with grid search. Despite these fluctuations, the R^2 remained stable at 0.92. The final model demonstrated a significant improvement in MAPE compared to the initial model, while the MAE and RMSE showed a slight increase in error.

Table 22 Approach B: Model Development (5-fold CV)

Model & Model Development Stage	**MAPE**	**MAE**	**RMSE**	R^2
Catboost				
Initial model and all features	21.77	421,813	647,108	0.92
Feature selection	21.44	429,666	660,020	0.92
Random search hyperparameter tuning	18.67	428,995	677,393	0.91
Grid search hyperparameter tuning	**18.35**	**422,314**	**665,811**	**0.92**
XGBoost				
Initial model and all features	22.04	436,544	671,285	0.91
Feature selection	21.84	441,744	679,921	0.91
Random search hyperparameter tuning	20.32	427,497	660,792	0.92
Grid search hyperparameter tuning	19.99	427,745	664,809	0.92
Random Forest				
Initial model and all features	21.84	451,157	698,535	0.91
Feature selection	21.82	452,722	698,729	0.91
Random search hyperparameter tuning	21.05	449,444	696,134	0.91
Grid search hyperparameter tuning	21.14	449,315	695,449	0.91

The XGBoost model with all features produced an initial MAPE of 22.04%, which is slightly higher than the Catboost model. The other metric also performed slightly worse compared to Catboost model. Feature selection resulted in small reduction in MAPE to 21.84%, but the other metrics exhibit minor increase in error. Random search hyperparameter tuning lowered the MAPE to 20.32% and grid search further improved this to 19.99%. However, similar to the Catboost model, the MAE and RMSE fluctuated and the R^2 remained constant at 0.92. The final model demonstrated improved performance in all metrics compared to the initial model with all features, albeit slightly worse than the final Catboost model.

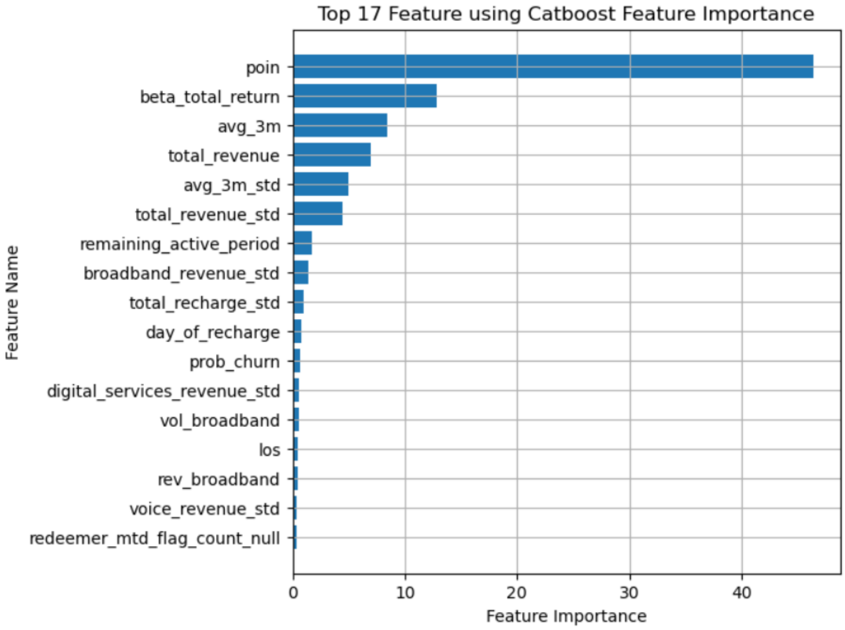
The Random Forest model initially achieved competitive results in MAPE at 21.84% compared to the other models but exhibited higher MAE and RMSE values. However, throughout the model development stage, the model performance increment is minimal. The hyperparameter tuning only able to slightly improve the metric, where the final model has MAPE value at 21.14%, the MAE at 449,315 and RMSE at 695,449. While the final model improved performance across all metrics compared to the initial model, it still performed the worst among the three models.

Comparing the three models, both Catboost and XGBoost consistently outperformed the Random Forest model in all metrics, with Catboost achieving the lowest final MAPE of 18.35% after grid search hyperparameter tuning. XGBoost demonstrated slightly better performance in MAE and RMSE compared to Catboost, albeit with marginal differences. However, based on the overall performance and the significant improvement in MAPE, Catboost was selected as the final model, as it exhibited an adequate performance gap while only slightly underperforming in MAE and RMSE.

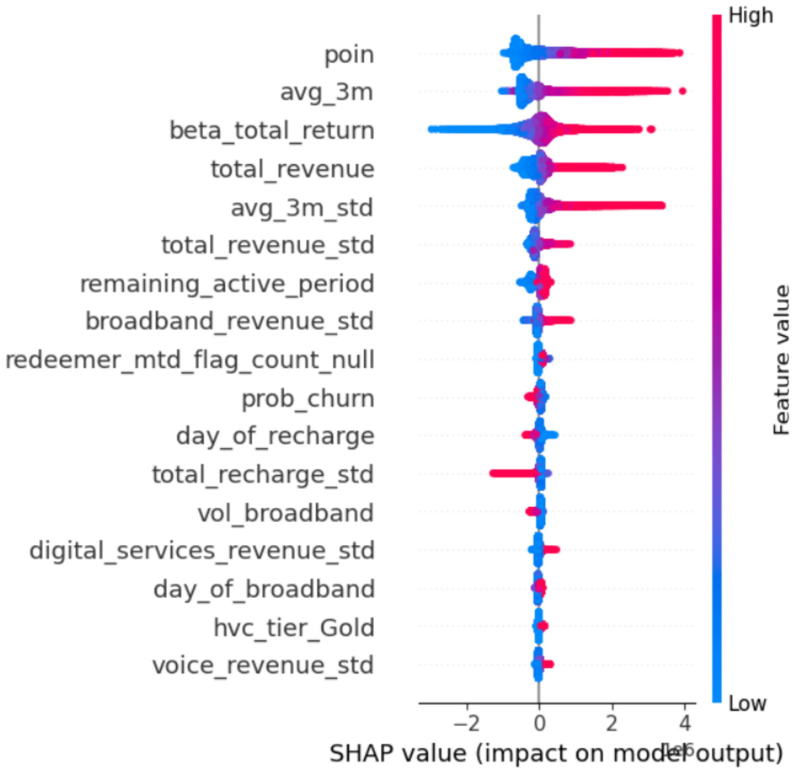
5.6.2 Model Interpretation and Feature Importance Analysis

5.6.2.1 Global Explanation

This approach, unlike the others, incorporates the volatility of revenue from customers by considering the beta value as a measure of risk in RAR calculation. Although this approach differs significantly from the others, the most important feature for the model remains consistent, which is "poin". In terms of feature importance, "poin" dominates with a value nearly four times that of the second most important feature. Similarly, in the SHAP value analysis, "poin" retains its position as the most important feature, but without any significant difference in its impact. "Beta_total_return," which indicates the beta value for the total return from the customer, is considered the second most important feature based on feature importance and the third most important feature according to SHAP value. While "prob_churn" is still among the top 10 most important features in this approach, its rank and impact are not as high as in the other approach. The detailed result is shown in Figure 40.



(a)

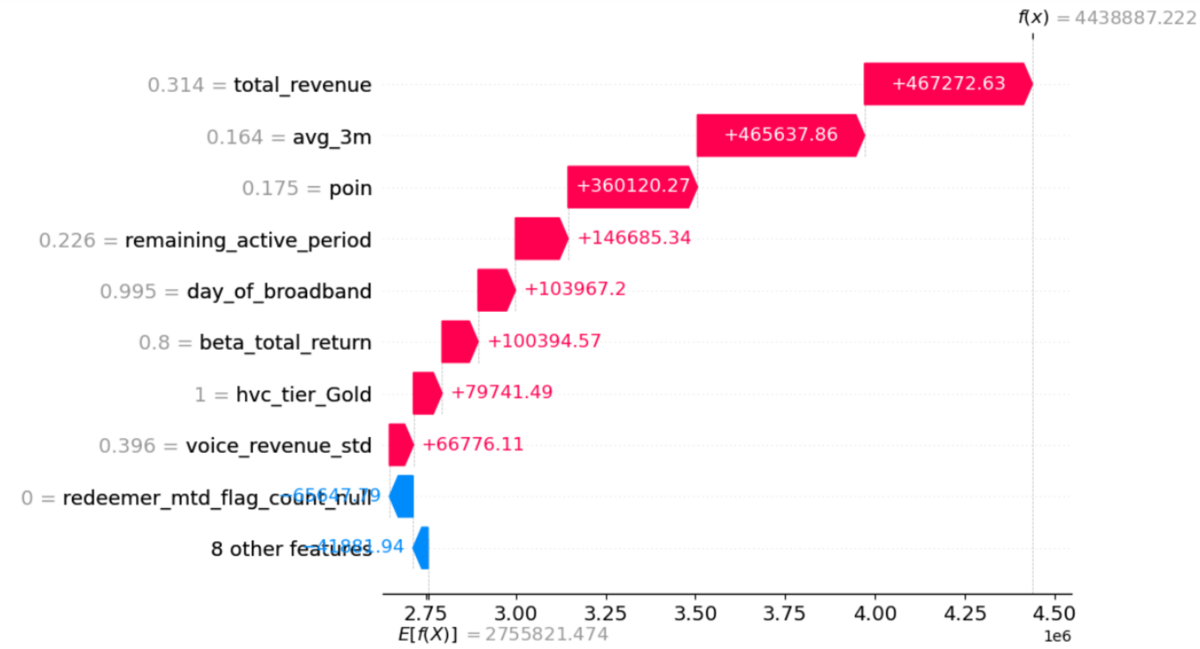


(b)

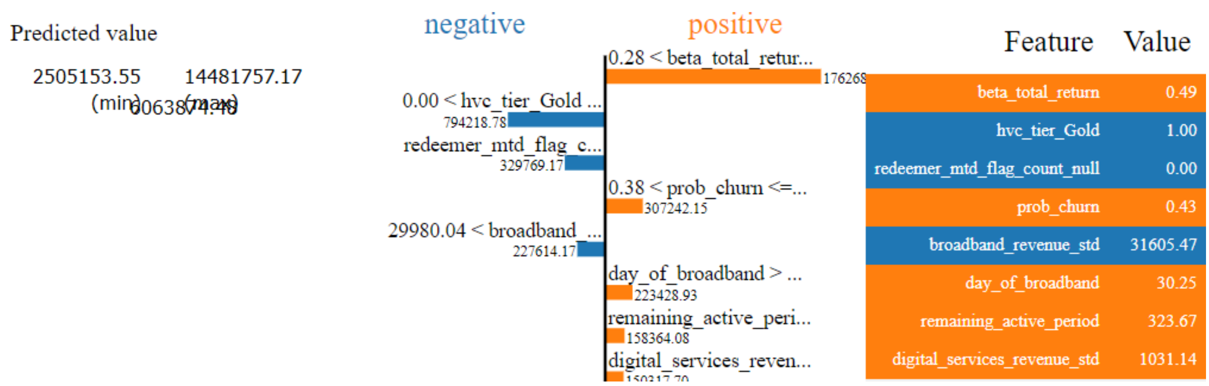
Figure 40 Approach B: (a) Feature Importance and (b) SHAP value

5.6.2.2 Local Explanation

Further evaluation using local explanation on the same data instance as the other approach reveals the impactful features according to SHAP value On the Figure 41, "total_revenue" and "avg_3m" are shown to have a positive impact, increasing the predicted result with similar SHAP value by 467,272 and 465,637 respectively. "beta_total_return" has a positive impact of 100,394, while "prob_churn" does not rank among the top 10 feature importances for this specific data point. On the other hand, according to LIME, the most important feature for this instance is "beta_total_return," which has a significant positive contribution of 1,762,681 on the predicted value. This value is more than double the value of the second most important feature, "hvc_tier_Gold," which has a negative LIME value of 794,218. Interestingly, "prob_churn" is also among the top 5 feature importances according to LIME, with a positive contribution of 307,242.



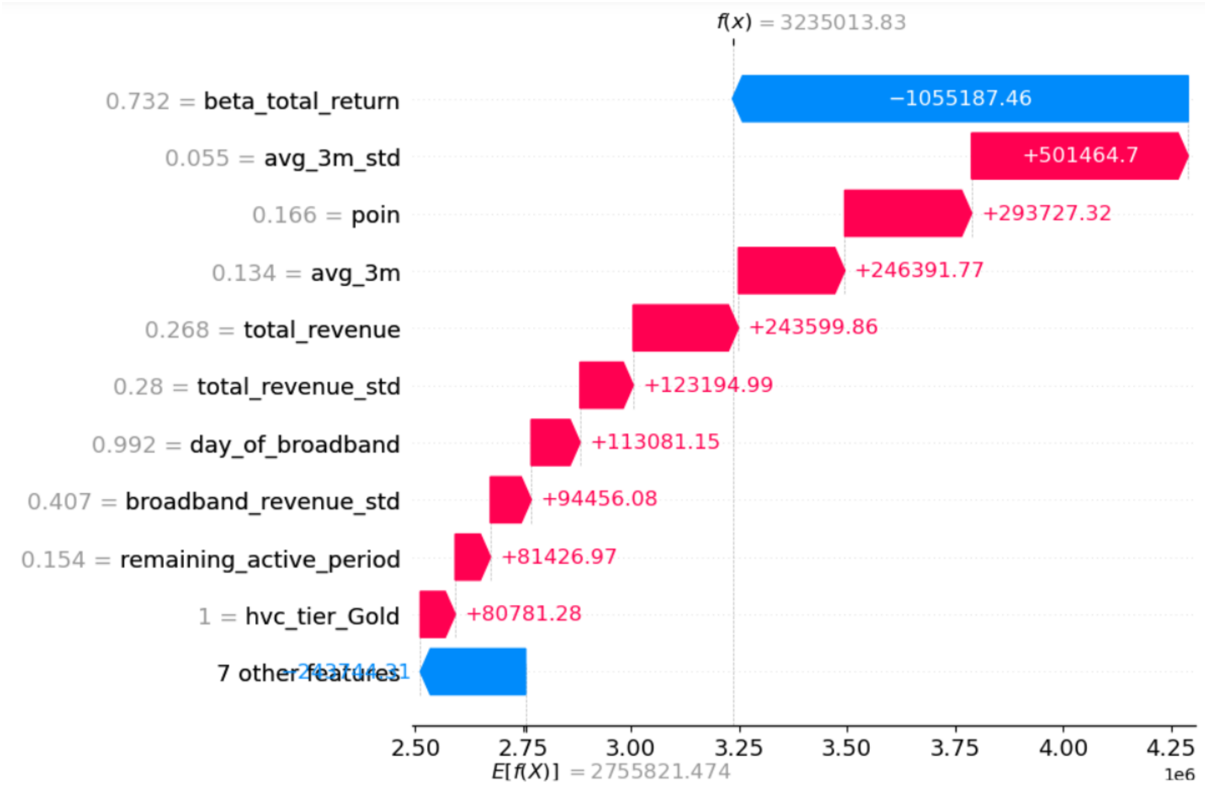
(a)



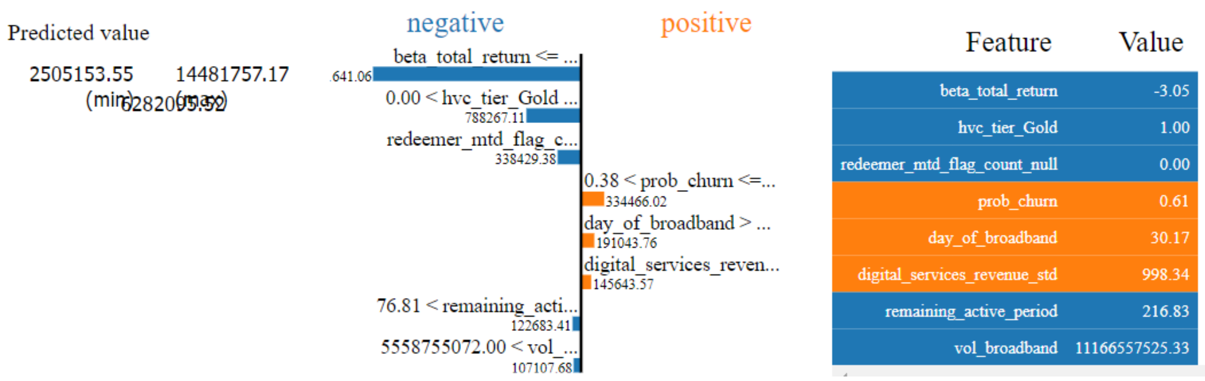
(b)

Figure 41 Approach B: (a) SHAP value and (b) LIME value for data sample X

For comparative purposes, an additional data point with a negative Beta value was selected, denoted as sample Z. The SHAP and LIME analyses were performed on this sample, and the results are presented in Figure 42. The feature "beta_total_return" emerged as the most influential feature in both SHAP and LIME explanations. According to the SHAP analysis, the negative Beta value had a significant negative contribution of -1,055,187 to the predicted result. Similarly, the LIME analysis highlighted the negative impact of the Beta value, with a high negative contribution of 3,001,641. These findings align with the observations presented in Subsection 5.2.2, which investigated the correlation between the Beta value and the calculated RAR. As established, the Beta value has a positive correlation with the RAR value. Therefore, in the case of sample Z, characterized by a negative Beta value of -3.05, the calculated RAR value is smaller than the actual total revenue. Consequently, the negative Beta value has a negative contribution to the predicted result.



(a)



(b)

Figure 42 Approach B: (a) SHAP value and (b) LIME value for data sample Z

5.6.2.3 Dependence Plot

Dependence plot analysis was conducted using SHAP to explore the interaction between the features and their corresponding SHAP values in Approach B. The top five most important features, according to the SHAP global explanation, were found to be consistent with Approach A: "poin," "avg_3m," "beta_total_return," "total_revenue," and "avg_3m_std." Therefore, the main feature selected for analysis was "beta_total_return," with the other top five features serving as the secondary features. The dependence plots illustrating these interactions are presented in Figure 43.

The findings from the dependence plots support the notion that the Beta value exhibits a positive correlation with the SHAP value. As depicted in the plots, higher Beta values are associated with higher SHAP values. However, it is important to note that the SHAP value differs from the RAR value. While the SHAP value tends to converge to zero as the Beta value approaches zero, positive Beta values do not necessarily result in positive SHAP values. In contrast, positive Beta values consistently yield higher SHAP values than the actual revenue. Interestingly, the interaction between the Beta value and the other features resembles the interaction observed between the probability of churn in the previous approach. Specifically, instances with high Beta values and high "poin" values tend to exhibit higher SHAP values, while those with low Beta values and low "poin" values tend to have lower SHAP values. However, unlike the previous approach, there appears to be a smaller amount of data with high values for the secondary feature. This suggests that the high values of the secondary feature are concentrated within a narrower range of Beta values.

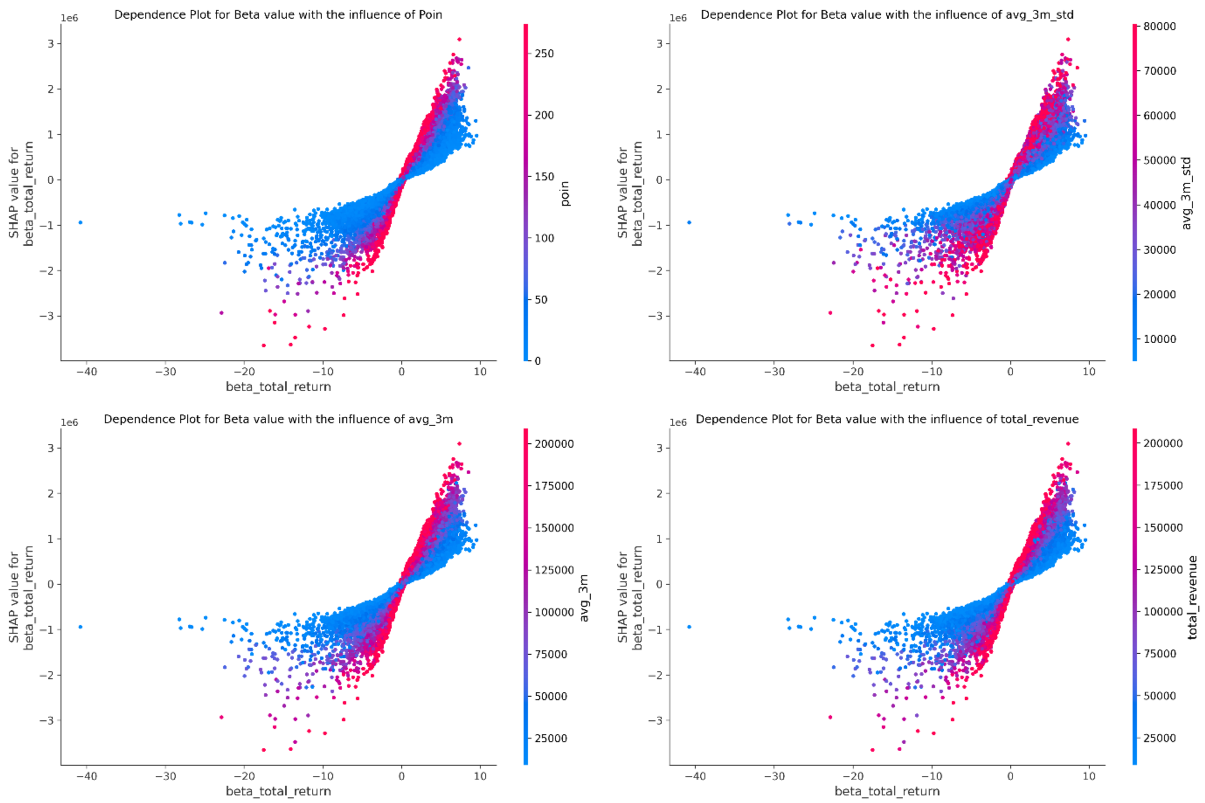


Figure 43 Approach B: SHAP Dependence plot of Probability of Churn

5.6.3 Model Validation

The best model from approach B then validated using different splitting strategies to assess its performance. The splitting ratio ranges from 50:50 to 90:10 and the results presented in the Table 23. The result indicates the stable performance across splitting ratio, where the best performance is achieved by 80:20 splitting ratio. These findings indicate that the model exhibits robustness and stability across different split ratios, demonstrating its ability to generalize well to unseen data.

Table 23 Approach B: Final Model Performances Across Different Split Strategies

Split Ratio	MAPE	MAE	RMSE	R^2
50:50	18.84	429,117	675,870	0.91
60:40	18.41	427,052	672,814	0.91
70:30	18.32	422,827	666,257	0.91
80:20	18.32	421,986	665,482	0.91
90:10	18.45	423,016	672,573	0.91

Five-fold CV results for Approach B are displayed in Table 24. The finding is consistent with Approach A, where both the training and test sets present a reasonable level of consistency across all the performance metrics through the five folds. The model's ability to perform uniformly well on unseen data and the training data indicates the effective prevention of overfitting. The minor fluctuations across the different folds in the performance metrics further imply the model's stability. Notably, the model's performance is not significantly influenced by any specific subset of the data, underlining the model's robustness and generalizability.

Table 24 Approach B: Final Model 5-fold CV Results

Fold	Train				Test			
	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE	R^2
1	14.75	336,476	551,978	0.9	18.41	421,542	663,835	0.9
2	14.68	335,076	550,289	0.9	18.29	423,410	667,693	0.9
3	14.77	337,057	553,040	0.9	18.08	419,802	661,276	0.9
4	14.65	337,731	555,873	0.9	18.46	421,160	661,286	0.9
5	14.71	335,086	549,399	0.9	18.5	425,654	674,963	0.9
Avg	**14.71**	**336,285**	**552,116**	**0.90**	**18.35**	**422,314**	**665,811**	**0.90**

Despite employing different approaches, the group split analysis in Approach B demonstrates a consistent pattern similar to the other approaches (Table 25). Test_group 3 consistently exhibits the lowest MAPE, while test_group 4 consistently shows the highest

MAE and RMSE. However, it is important to note that the performance of the remaining groups remains stable across all metrics. This consistent pattern reinforces our confidence in the stability and reliability of the best model across different splitting methods.

Table 25 Approach B: Final Model Group Split Results

Test	Train				Test			
group	MAPE	MAE	RMSE	R^2	MAPE	MAE	RMSE	R^2
1	14.76	328,323	541,199	0.94	18.42	493,205	751,950	0.91
2	14.31	334,576	551,253	0.94	18.29	402,007	638,864	0.91
3	14.09	337,393	554,664	0.94	21.66	408,716	649,884	0.91
4	14.89	333,949	546,087	0.94	18.46	624,200	897,520	0.89
Avg	**14.51**	**333,560**	**548,301**	**0.94**	**19.21**	**482,032**	**734,555**	**0.91**

Learning curve analysis also performed on this approach to further investigate the learning efficiency and model robustness as showed in Figure 44. The finding is also similar with Approach A, where the training score increase with the size of the training set while the validation scores decrease. Furthermore, the gap between the training and validation score became smaller as the training size increases. This pattern indicate that the model is not overfitting as it is generalizing well to unseen data.

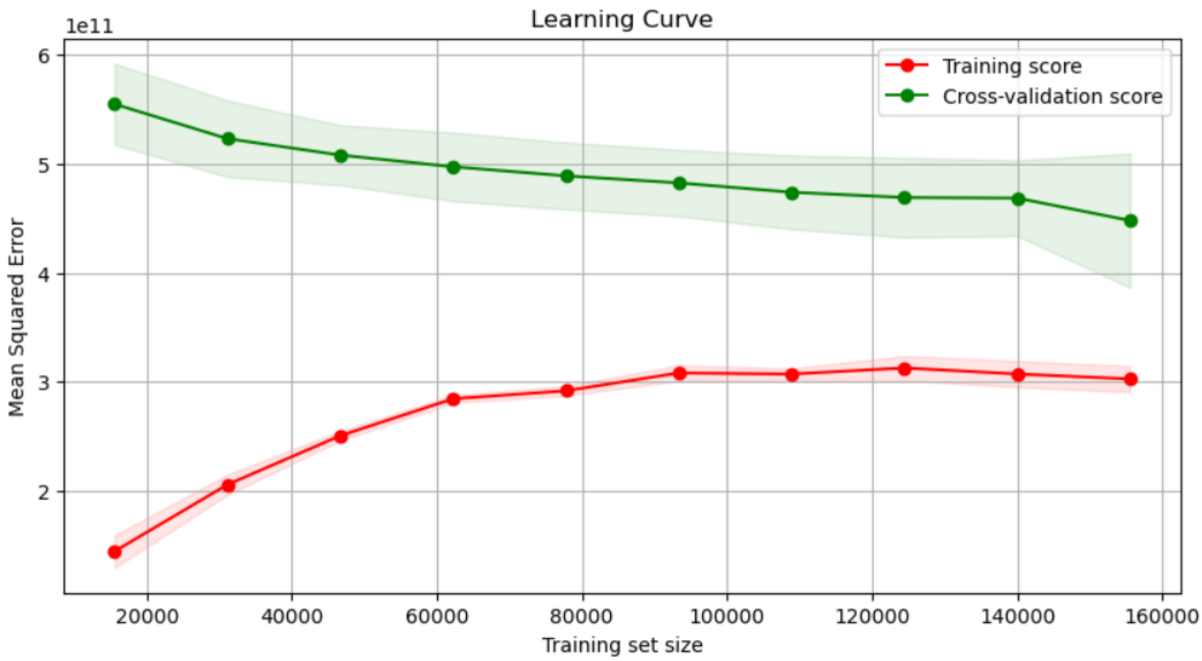


Figure 44 Approach B: Final Model Learning Curve Analysis

Additional validation of the model's output involves a comparative examination between impactful features determined by the ML model and variables stressed in conventional calculation methods. Approach B employs a distinct RAR calculation formula outlined in Equation 4.2. In this formula, the discount rate equals the expected return from all customers, multiplied by the customer's Beta value. As such, the customer's Beta value emerges as a critical feature in traditional RAR calculation. Consistently, the final model also identifies the Beta value as a top-tier feature. It ranks within the top three features in the Global Explanation using SHAP and feature importance and persistently appears among top features in Local Explanation using SHAP and LIME. Furthermore, from Figure 42's dependence plot, a positive correlation is observable between the Beta Value and the SHAP value, with higher Beta values corresponding to lower SHAP values. This correlation aligns with traditional RAR calculation methods, specifically for this case. Given that the expected return from all customers is -1.3%, a negative Beta value will result in a positive discount rate, while a positive Beta value will yield a negative discount rate. And the positive discount rate will result a lower calculated RAR value and negative discount rate will result in higher RAR value than the actual revenue.

The other top important features also signify the alignment with traditional calculation methods. The features like "poin", "avg_3m_std", "avg_3m", "total_revenue", and "total_revenue_std" also have been identified as the top important features. By focusing on these features, which represent various aspects of customer spending and derived revenue, the model stays consistent with the RAR calculation formula (Equation 3.1). Interestingly, features like "avg_3m_std" and "total_revenue_std" capture the volatility of customer spending, a natural occurrence over the six semesters of data collection. The inclusion of this volatility in the RAR calculation further proves the model's thoroughness and its alignment with conventional methods.

5.7 Summary of Findings

The following section presents a detailed summary of the primary findings from the analysis, focusing on model predictive power, insights from XAI analysis and the relevance with traditional RAR calculation.

1. Predictive Power of Models
    a. The churn model operates with nearly the same performance when utilizing only the top 17 features as when incorporating all features, representing an approximate decrease in performance of only around 0.1% across all ML algorithms examined in this study.

   b. Among the various models tested for churn prediction, the XGBoost classifier demonstrated superior accuracy and F1 score, outperforming other alternatives.
   c. The best performing model for predicting RAR value across all approaches was found to be the CatBoost regressor. The XGBoost also displayed competitive results and Random Forest consistently underperformed relative to the others.
   d. Comparing the model's performance with top 150 features versus top 17-18 features for RAR prediction reveals negligible differences, with MAPE worsening by less than 0.5%.
2. XAI analysis
   a. Across all approaches to RAR calculation, the feature "poin" consistently emerged as the most significant factor in Global explanations using both SHAP values and Feature Importance.
   b. For Approach A, the probability of churn surfaced as the second most crucial feature using Global Explanation with both SHAP value and Feature Importance. It often ranked as the most important feature in selected samples in Local Explanation using LIME and SHAP.
   c. In Approach B, beta_value stands out as the second most essential feature in Global Explanation, using both SHAP and Feature Importance, and often ranking as the top feature in Local Explanation.
   d. Probability of churn still ranks among the top features in Global Explanation for Approach B.
   e. Interaction patterns were observed between the churn probability, beta_value, and other top features, impacting the SHAP values in distinct ways.
3. Relevance with traditional model
   a. A negative correlation exists between the probability of customer churn and the predicted RAR value. This correlation is aligned with the traditional calculation method.
   b. Beta value in approach B shows a positive correlation with the predicted RAR value. This correlation is also aligned with the traditional calculation method.
   c. Both the probability of churn in Approach A and the beta_value in Approach B, which rank as highly significant features, are in line with the traditional RAR calculations employed in this study.

d. Similarly, the features avg_3m, avg_3m_std, total_revenue, total_revenue_std consistently appear among the top features for both approaches A and B, further confirming their relevance in RAR calculation.

These findings collectively provide a deeper understanding of the factors influencing churn predictions and RAR calculations, thereby facilitating the creation of more accurate and robust models for these purposes.

6 Conclusion

This research presents a significant step forward in advancing the understanding of customers' value within the telecommunications industry. It introduces and implements a novel approach to assessing customers' value through the lens of Risk-Adjusted Revenue (RAR), offering a more comprehensive perspective on customer worth that takes into account not only potential revenue but also associated risks. To the best of our knowledge, the uniqueness of this work lies in its specific application to the telecommunications sector, marking the first time that RAR has been employed within this industry. This contributes to filling a clear research gap and broadens the scope of existing studies on RAR, which have previously been concentrated in other sectors.

Moreover, the practical implications of this research are substantial. By providing telecommunications companies with a more nuanced understanding of their customers' value, the developed models allow for data-backed, tailored customer treatments and portfolio management. This capacity to better identify and understand customers could enable companies to optimize their profits and drive their business strategies in a more informed and efficient manner.

Overall, the present research not only deepens the academic understanding of RAR but also demonstrates its potential as a valuable tool for customers' value assessment in the telecommunications industry. As such, it can act as a springboard for further research and innovative practices in both the academic and industrial fields.

6.1 Answer to Research Questions

This study set out with the primary research question of identifying effective ways to apply ML techniques for predicting risk-adjusted CLV in the non-contractual (B2C) setting of the telecommunications industry. To fully address this central inquiry, the research was segmented into a series of sub-questions designed to holistically understand the current literature, identify gaps, and then propose, develop, and evaluate an innovative solution. The methodology employed was a combination of the DSRM as the main methodology on the study and the CRISP-DM as the guidance to build the ML model. This combination ensures a comprehensive approach to both the conceptual and practical aspects of the study. The primary findings suggest a robust and practical ML model that successfully predicts risk-adjusted CLV, that represented by RAR, while incorporating significant aspects of risk from the probability of customers' churn and the volatility of revenue. The subsequent paragraphs will delve into a detailed discussion on how each sub-question was addressed and how the outcome contributed to answering the primary research question.

SQ1. How has the incorporation of customers' risk into CLV calculation evolved over time in the literature?

The evolution of the incorporation of customers' risk into CLV calculations over time has been largely influenced by the financial portfolio theory. Early efforts mainly drew from MPT and the CAPM, treating customers as risky assets with distinct risk profiles that needed management for optimal returns (Dhar and Glazer, 2003; Ryals, 2003). However, researchers highlighted CAPM's disregard for unsystematic risk, prompting a shift towards Markowitz's PST as a more comprehensive approach. This shift was showed by applications of PST in diverse contexts, such as (Sackmann et al., 2010) and B2B settings (Tarasi et al., 2011; Juhl and Christensen, 2013), and further extended through hybrid models integrating stochastic CLV modelling and ex-ante customer portfolio optimization (Norouzi and Albadvi, 2016). Recognizing limitations in models considering a single type of risk, the MSR approach emerged (Singh et al., 2013; Singh and Singh, 2016; Machado and Karray, 2022a). This approach considers various risk factors pertinent to specific industry and business settings, providing a more comprehensive understanding of customers' risk. For example, the specific risk associated with customers varies widely between industries, necessitating distinct risk-adjusted CLV metrics for each. With technological advancements, particularly the rise of AI-enhanced predictive analytics powered by real-time data, further enhance the driver of the need for sophisticated risk-adjusted CLV models in the future. In conclusion, the journey of incorporating customers' risk in CLV calculation has been marked by evolution and adaptation, driven by the need for more precise, comprehensive, and context-sensitive models that more accurately capture and reflect the multifaceted nature of customers' risk. Further insights on this evolution can be explored in Subsection 2.3.2.

SQ2. What are the industries or domains where customers' risk has been incorporated into CLV calculation?

Incorporation of customers' risk into CLV calculations has been observed across several industries, mainly clustered into three categories: FSI, B2C (non-FSI sectors), and B2B settings. The majority of studies (44%) were conducted in the FSI, spanning diverse sectors such as insurance, credit card services, and peer-to-peer lending. B2C sectors outside the FSI constitute the second largest group (30%), with industries ranging from e-commerce and pharmacy to airlines. B2B settings make up the remaining studies (26%), covering various sectors like medical instrument providers and commodity companies. A detailed analysis can be found in Subsection 2.3.3.

SQ3. What are the commonly used methods for incorporating customers' risk into the calculation of customers' value in the industry?

The methods for incorporating customers' risk into customers' value calculations depend on the specific application area as detailed in Subsection 2.3.4. Nevertheless, some common techniques are evident:

- **Mean-Variance Analysis**: Predominantly used in financial portfolio concepts for customer portfolio optimization, this approach segments customers based on demographic features before predicting the optimal portfolio composition (e.g., Buhl & Heinrich, 2008; Homburg et al., 2009; Tarasi et al., 2011; Sackmann et al., 2010; Viviani et al., 2021; Machado & Karray, 2022b). Studies like Tarasi et al. (2011) and Viviani et al. (2021) leverage mean-variance analysis for optimizing portfolio composition, considering segment-specific variables and transition probabilities.
- **Direct Calculation & Predictive Modeling**: For certain applications, risk-adjusted CLV is calculated directly through mathematical modeling (e.g., Ryals and Knox, 2005), while others build predictive models, such as the scorecard predictive model by So et al. (2014), to identify more profitable customers.
- **Customer Segmentation & Modeling**: Some studies use customer segmentation followed by risk calculation. For instance, Albadvi and Norouzi (2013) employ RFM segmentation, followed by Pareto/NBD modeling for risk-adjusted CLV.
- **Data Envelopment Analysis (DEA)**: DEA is used to compute measures of risk-adjusted revenue, as demonstrated by Singh et al. (2013) and Singh and Singh (2016) in their respective studies.

SQ4. What is the state-of-the-art ML model used to predict the risk-adjusted CLV in the industry?

The state-of-the-art ML model for predicting risk-adjusted CLV in the industry as discussed in Subsection 2.3.5 is a hybrid ML framework, proposed by Machado & Karray (2022a). They predict the RAR of customer in p2p lending by leveraging the strengths of both supervised and unsupervised learning to efficiently handle high-dimensional datasets and produce more robust predictions. In this framework, unsupervised learning (clustering algorithms like k-Means++, k-Means random, DBSCAN) is first employed to segment customer data, which is then used as a variable in supervised learning models (predictive algorithms like Adaptive Boosting, Gradient Boosting, Decision Tree, Random Forest, Support Vector Machine, and Artificial Neural Network) to estimate the RAR value. Among various combinations tested, Machado & Karray found that combinations of Decision Tree, Random Forest, or Gradient Boosting with k-Means++ outperform standalone models in predictive power and processing time. Hence, the hybrid ML approach presents significant potential for RAR prediction in the industry, offering vital guidance for future studies.

SQ5. What is the most significant customer's type of risks to be considered when assessing customers' value in the Telecommunication industry?

The SLR study identified nine different types of customers' risks, with the most significant risks in the previous studies being:

- **Volatility of Customers' Income**: This risk indicates the variability or uncertainty in a customer's income, which can directly impact their spending behavior. It's a readily measurable risk that can be incorporated into CLV calculations. It is calculated using the variance in customer transactions over time. Different studies have used slightly different formulae, but all aim to capture this volatility. Wangenheim and Lentz (2005), Homburg et al. (2009), Singh et al. (2013), and Machado & Karray (2022a) are among the studies that have considered this type of risk.
- **Beta Risk**: This risk comes from financial portfolio theory (CAPM and MPT) and measures the volatility of an asset's price fluctuations compared to a benchmark. It is often used to calculate the discount rate in CLV calculations. It has been incorporated in the studies by Wangenheim & Lentz (2005), Buhl & Heinrich (2008), Tarasi et al. (2011), Albadvi & Norouzi (2013), Singh et al. (2013), and Machado & Karray (2022b).
- **Customer Churn**: This refers to the risk of a customer discontinuing their relationship with the company. It is now considered a significant risk for many companies, including those in the Telecommunication industry. Singh and Singh (2016), Hai-wei et al. (2006), Ryals (2003), and Ryals and Knox (2005) have all emphasized the importance of considering this risk in CLV calculations.

Other industry-specific risks were also identified, such as PD in the financial services industry and claim risk in the insurance industry. However, the three above-mentioned risks are the most significant when assessing customers' value in the previous studies. The detailed discussion on this topic can be found in Subsection 2.3.6.

SQ6. How much historical data and what time period should be considered to evaluate the risk-adjusted customers' value in the industry?

The time period and amount of historical data used for evaluating risk-adjusted customers' value differ greatly based on the industry and the method of analysis. The shortest observation period was 39 weeks (Sackmann et al., 2010) and the longest was 10 years (Buhl & Heinrich, 2008). For the telecommunications industry specifically, a quarter of a year was considered appropriate (Homburg et al., 2009). On average, B2C sector studies used about 2.85 years of data, B2B used 3.5 years, and FSI used 4.9 years. The number of customers observed also varied greatly, from as few as 10 in one study (Ryals & Knox, 2007) to as many as 2 million in another (Machado & Karray, 2022a, 2022b). The methodology also impacts the required data volume. For instance, ML algorithms require larger datasets for accuracy. A comprehensive analysis is available in Subsection 2.3.7.

SQ7. What specific risks identified from the literature are most relevant to the telecommunication industry, and how can these be quantified for inclusion in the ML model?

In order to address the question of which specific risks, identified from the literature, are most relevant to the telecommunication industry and how these can be quantified for inclusion in the ML model, this study adapts and incorporates two distinct risks: the probability of customer churn and the volatility of revenue from customer that represented by beta value.

The selection of the probability of customer churn as a risk factor is influenced by its high occurrence in the telecommunications industry, particularly amongst prepaid customers who are able to switch providers at any given time. To quantify this risk for inclusion in the ML model, the approach proposed by Ryals & Knox (2005) is followed. In this method, the customer's probability of churn is integrated into the formula for the discount rate, with an assumption made that the dataset used in the study is representative of the entire market. This churn probability is estimated using a ML model developed following the CRISP-DM framework and the ML pipeline, which presented in Section 5.1.

The volatility of revenue from each customer is selected as the second risk due to its prevalence as a common risk across industries. The quantification of this risk uses the Beta value, which assesses the volatility of revenue from each customer compared to the volatility of the market, following the CAPM theory as adopted by Dhar & Glazer (2003), Buhl & Heinrich (2008), and Machado & Karray (2022). This approach captures the variability and fluctuations in the customer's revenue contributions, thereby providing a more comprehensive and robust assessment of customers' risk.

Incorporating these two risks into the calculation of RAR value makes the metric more comprehensive and facilitates meaningful comparisons across experiments. It also provides a robust foundation for evaluating the financial impact of various risks on customer revenue, enabling informed decision-making and resource allocation. The method used to calculate RAR builds upon the base model introduced by Berger and Nasr (1998) but includes adaptations to the discount rate to account for the risk associated with each customer, which described in more detailed in Section 4.2.

SQ8. How to develop a ML Model to predict the risk-adjusted CLV in telecommunication industry?

The development of ML model tailored for the telecommunication sector to predict the risk-adjusted CLV, adhering to the structured approach of the CRISP-DM methodology and as illustrated in Figure 12. The pipeline integrated initial features that delved deep into customer behavior. These features captured transaction histories, service utilization patterns—spanning voice, broadband, and more—recharge behaviors, and device usage details. Monthly data underwent rigorous processing: missing values were addressed, categories streamlined, and new attributes introduced. Following this, the data was aggregated, and additional features were engineered, involving the creation of novel attributes and outlier management. Categorical variables were then encoded using OHE, and features were selected based on multicollinearity and correlation analyses, complemented by feature importance evaluation. Given the intricate dynamics and vastness of telecom data, algorithms such as CatBoost, XGBoost, and RF were selected. The dataset was scaled using a MinMax scaler, followed by hyperparameter optimization to improve the model's predictive capabilities. To ascertain reliability, the best-performing model underwent validation using diverse data split methods and cross-validation. A spotlight was placed on feature importance analysis and XAI techniques, emphasizing the role of telecom-specific attributes in determining risk-adjusted CLV. Finally, the most influential features derived from the model were compared with variables used in traditional risk-adjusted CLV computations. An in-depth exposition of the ML model development technique is presented in Chapter 4, particularly in Sections 4.3 through 4.7.

SQ9. How does the strategy of splitting data into training and test sets affect the performance of the ML models for risk-adjusted CLV prediction?

The analysis includes various strategies for splitting data to ensure the model's robustness. Train-test split ratios from 50:50 to 90:10 was examined. Across these various ratios, the model performance displayed minimal fluctuation, indicating the model's stability.

A 5-fold Cross-Validation was also implemented. The ML model showed stable performance across both training and testing sets, demonstrating no signs of overfitting. Stability was further indicated by the insignificant fluctuation in performance metrics across different folds, indicating the model's performance isn't heavily reliant on any particular data subset. This demonstrates the model's generalizability and robustness.

The model was further scrutinized using a group split strategy, with "region_tier" serving as the differentiating factor. Even though there were slight variations among the group splits, the model's performance remained consistently strong, emphasizing its robustness.

Learning curve analysis was also conducted, using percentages of data ranging from 10% to 80%. The trends indicated that as the training data volume increased, the model's ability to fit the training data decreased, while its generalizability to unseen data improved. The small difference between the training and validation scores at larger training sizes indicates that the model is not overfitting and is capable of generalizing to unseen data.

For performance evaluation, a combination of MAPE, MAE, RMSE, and R^2 metrics was used. These metrics give a comprehensive understanding of the model's accuracy, error rate, and its ability to explain variance in the data. The consistently strong performance across these metrics supports the model's robustness.

In conclusion, the strategies employed for data splitting significantly contribute to determining the reliability and robustness of ML models for predicting risk-adjusted CLV. The model in this study exhibits stable performance, robust generalizability, and a significant ability to handle varying amounts and structures of data, validating its suitability for predicting risk-adjusted CLV in the telecommunications industry. The detailed result for each approach could be found in Model validation subsection in Chapter 5.

SQ10. What is the most important feature/variable to predict the risk-adjusted CLV in telecommunication industry?

The ML model identified several key features as the most important predictors of risk-adjusted CLV in the telecommunications industry. These are primarily the customers' loyalty point (poin), the average revenue in the last 3 months (avg_3m), including the standard deviation (avg_3m_std), and the total revenue (total_revenue) including its standard deviation (total_revenue_std). Notably, different approaches yielded a few additional significant features. In approach A, the probability of customers' churn (prob_churn) emerged as an important predictor. This reflects the critical role customer retention plays in determining CLV, especially in the telecommunications industry where churn rates can significantly impact revenue. Meanwhile, in approach B, the customers' beta value (beta_value) was identified as a significant feature. This illustrates the model's recognition of customers' risk and its relation to overall customers' value. This is particularly crucial in a risk-adjusted CLV calculation, where customers' value is adjusted according to their associated risk.

SQ11. How does the most important feature from the model compared to the traditional calculations' method?

The most important features extracted from the ML model align closely with key variables utilized in traditional calculation methods, thus affirming the model's validity. The top five features: "point", "avg_3m", "avg_3m_std", "total_revenue", and "total_revenue_std", essentially represent customer spending or revenue – a fundamental element in traditional CLV calculations. Interestingly, for Approach A, the ML model identified "prob_churn" as one of the top five critical features, mirroring its importance in traditional methods for this approach. Similarly, for Approach B, the model highlighted "beta_value" as a significant factor, aligning it with this approach's traditional methods.

Furthermore, the model successfully captured that customer spending volatility substantially influences the RAR value, as inferred from the importance of the standard deviation of average revenue from the last 3 months and the total reveue. These features represent the spending variability during the observation period. Moreover, the final model's correlation between "prob_churn" and "beta_value" with the predicted RAR value is consistent with the correlations observed in traditional calculation methods. In this correlation, "prob_churn" exhibits a negative relationship with RAR, while "beta_value" has a positive correlation, further validating the model's accuracy and relevance in predicting RAR value.

6.2 Implications

6.2.1 Theoretical Implications

The implications of this research for theoretical advancements are significant, offering novel insights and methodologies in the domain of the telecommunications industry.

Firstly, this study pioneers a research focus on calculating and predicting RAR in the telecommunications industry. This provides an original perspective that enriches the broader discourse on RAR and contributes to the body of knowledge by situating these processes within the unique context of the telecommunications sector. Secondly, this research contributes to the existing literature on integrating risk in customer valuation. It introduces two novel methods for calculating RAR, expanding the methodological toolbox available to researchers and practitioners alike in this field. This is a substantial step forward in understanding and quantifying customer-related risks in this industry.

Thirdly, the application of ML for RAR prediction in this study marks a significant development. Only a handful of studies have fully utilized ML to predict RAR. Therefore, the use of ML in the study not only substantiates the potential of ML in this context but also demonstrates its practical application. This application strengthens the case for incorporating ML techniques in future research on RAR prediction. Lastly, this study has significant implications for the field of XAI. This research is the first to leverage XAI to assess feature importance in RAR modeling and compare its results with traditional calculation methods. This approach highlights its potential for delivering clearer, more understandable ML outcomes and sets a precedent for future research.

6.2.2 Practical Implications

The practical implications of the result of this study are manifold and offer significant potential for a more informed and effective marketing strategy. Firstly, the models can serve as robust tools for implementing targeted customer retention campaigns. The churn model could help identify the customer probability of churn, hence enabling the marketing team to devise personalized retention strategies to mitigate this risk. The marketing to choose to give campaign to the customer with medium to high probability to churn, as they are easier to be roped to stay, while for the customer that have high risk of churn, the marketing team could devise a plan to whether improve their retention that might need more budget allocation or just let them go.

The RAR model, which incorporates churn risk, offers multifaceted advantages to the marketing team. It serves as a tool to identify valuable customers, thus presenting opportunities to not only safeguard but also grow the company's revenue. Tailoring marketing strategies to these high-RAR customers can help sustain and potentially increase their revenue contributions. Additionally, these models facilitate a more precise customer segmentation based on churn risk and RAR value. For instance, immediate attention and targeted marketing interventions may be necessary for high-RAR customers with a high risk of churn. In contrast, customers with a low churn risk but high RAR could be the center of strategies designed to further amplify their spending.

Moreover, the RAR model's integration of risk factors allows it to guide risk-informed decision-making within the marketing team, proving particularly beneficial when allocating budgets for retention programs. Customers posing a significant financial risk can be targeted more aggressively, enabling the marketing team to construct the most optimal customer portfolio according to their customer segments. Each segment, having unique needs, plans, budgets, and returns, calls for a tailored approach. In this way, the marketing team can devise strategies that maximize the trade-off between profit and retention efforts. In terms of product and service strategies, insights from the models regarding impactful features can inform targeted product recommendations and personalized offerings. This tailored approach can bolster customer satisfaction, thereby reducing churn and potentially amplifying RAR.

Lastly, the identification of key features affecting churn and RAR can provide valuable feedback for product development and management teams. This information can be leveraged to refine the company's offerings, increasing customer satisfaction, and decreasing the risk of churn. Overall, integrating the churn and RAR models into marketing strategy can lead to a more data-driven approach, more effective resource allocation, improved customer retention, and ultimately, growth in risk-adjusted revenue.

For the Company Specifically:

This research's insights offer a potent framework for enhancing the company's marketing strategies. When the churn and RAR models are integrated into the company's systems, it could provide the marketing team with a new dimension to assess customers' value. Feedback from company representatives highlighted the potential of these models in refining strategies for high-value customer retention and devising approaches for low-value customers. By capitalizing on this study's unique insights, the company can further sharpen its customer-centric strategies, optimize marketing and product offerings, and solidify its revenue growth in a competitive market.

6.3 Limitations

There are several limitations to the research that need to be considered. One of the most significant limitations is the selection of risk parameters. The study primarily focuses on the probability of customer churn and the Beta value as the source of customers' risk. However, broader risk factors could be explored, including revenue volatility, geographical variances, and the fluid dynamics of customer segmentation over the observation period. Additionally, this study estimates the RAR based on total revenue, overlooking the diverse range of revenue sources within this sector, including voice, SMS, broadband, and digital offerings. Therefore, a more granular analysis of these components might yield a richer understanding of RAR value in the telecommunication industry.

The limitations in this study stem from both the scope of data and the methodology adopted. Our analysis reveals that while the learning curve shows a notable enhancement in the model's predictive performance with an increasing sample size, the constraints on the size and duration of the available data inevitably impact the depth and accuracy of the insights gained. Furthermore, the study employs only three different ML algorithms for predicting the RAR value. The exploration of more sophisticated models, such as neural networks, could potentially provide a better understanding of customer behavior patterns. The optimization approach for hyperparameters, employing a combination of Random Search and Grid Search, also poses a limitation. Given the continuous nature of the hyperparameter values, the optimal value could potentially be missed, suggesting that a more sophisticated optimization method might yield superior results.

Finally, the method employed for evaluating feature importance in this research, using standard SHAP values, represents another limitation. The significant features in predicting the RAR, as highlighted by SHAP values, predominantly reflect different forms of customer spending such as "poin", "avg_3m", and "total_revenue". Also, there are "avg_3m_std" and "total_revenue_std", which correlate with the main features mentioned before. However, while SHAP values offer a robust measure of feature importance, they do not account for the potential causal structures in the data (Janzing et al., 2019). The limitations of SHAP analysis stem from its tendency to distribute attribution equally over features that provide similar information, thereby putting all features on an equal footing in model explanation. This could overlook the intricate interdependencies among features and a more nuanced interpretation might be required to discern their true impact on RAR.

6.4 Future Research

From the results and limitations of the study, there are several areas to explore for future research. First, examining other risk factors inherent in the telecommunications sector could significantly broaden the comprehension of customer behavior. Subsequent investigations might delve into a wider variety of risk factors, such as the shifting the customer segments due to the volatility of usage and exploring the risk associated with customer location. The probability of changing customer segments could be used as a method to capture the volatility of revenue from customers, which is inherent in all types of customers. On the other hand, the customer's location could be associated with the risk of competitor availability and network performance in the area. Furthermore, features related to customer location are regarded as an important feature in the customer churn probability model, hence deeper exploration could improve the understanding of the area.

Moreover, enhancing the methodology for computing the RAR presents another potential area for exploration. This study has primarily focused on overall revenue, therefore, breaking down different revenue components in future research could provide a more comprehensive perspective on the RAR. The revenue components could encompass various telecom services, including voice, SMS, broadband, and digital offerings, offering a broader RAR view across the industry's services. Another area for future research is the use of advanced techniques like neural networks. This approach may reveal deeper patterns in customer behavior, leading to improvements in the prediction and calculation of RAR. Similarly, exploring more advanced methods for hyperparameter optimization, such as Bayesian optimization, might enhance the model's performance.

Additionally, adopting advanced methods for assessing feature importance, such as causal and asymmetric SHAP values, that take into account potential interdependencies and causality within the data could provide a more detailed understanding of the effects of different features on RAR. In conclusion, these directions for future research could continue to enrich the understanding of RAR in the telecommunications industry and lead to more accurate and actionable insights.

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access, 6*, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *Sage Open, 9*(1). https://doi.org/https://doi.org/10.1177/2158244019829575

Albadvi, A., & Norouzi, A. (2013). Using downside CAPM theory to improve customer lifetime value prediction in non-contractual setting. *Management Science Letters, 3*, 3003-3012. https://doi.org/10.5267/j.msl.2013.10.021

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In M. Berry, A. Mohamed, & B. Yap, *Unsupervised and Semi-Supervised Learning* (pp. 3-21). Springer, Cham. https://doi.org/10.1007/978-3-030-22475-2_1

Al-Mashraie, M., Chung, S. H., & Jeon, H. W. (2020). Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach. *Computers & Industrial Engineering, 144*. https://doi.org/0.1016/j.cie.2020.106476

Alpaydin, E. (2010). *Introduction to Machine Learning.* The MIT Press.

Alpha Spread. (2023). Retrieved 06 09, 2023, from Alpha Spread: https://www.alphaspread.com/security/nyse/tlk/discount-rate

Annarelli, A., Battistella, C., Nonino, F., Parida, V., & Pessot, E. (2021). Literature review on digitalization capabilities: Co-citation analysis of antecedents, conceptualization and consequences. *Technological Forecasting and Social Change, 166*. https://doi.org/10.1016/j.techfore.2021.120635

Asadi, N., & Kazerooni, M. (2023). A stacked ensemble learning method for customer lifetime value prediction. *Kybernetes*. https://doi.org/10.1108/K-12-2022-1676

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies, 1*(1), 377-386. https://doi.org/10.1162/qss_a_00019

Benedek, G., Lublóy, Á., & Vastag, G. (2014). The Importance of Social Embeddedness: Churn Models at Mobile Providers. *Decis. Sci., 45*, 175-201. https://doi.org/10.1111/deci.12057

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artifcial Intelligence Review, 54*, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Berger, P., & Nasr, N. (1998). Customer lifetime value: marketing models and applications. *Journal of Interactive Marketing, 12*(1), 43-54.

Binh, T. V., Thy, N. G., & Phuong, H. T. (2021). Measure of CLV Toward Market Segmentation Approach in the Telecommunication Sector (Vietnam). *SAGE Open, 11*(2), 1-9. https://doi.org/10.1177/21582440211021584

Borle, S., Singh, S., & Jain, D. (2008). Customer Lifetime Value Measurement. *Manag. Sci., 54*, 100-112. https://doi.org/10.1287/mnsc.1070.0746

BPS Indonesia. (2022). *Jumlah Penduduk Menurut Kelompok Umur dan Jenis Kelamin, 2022*. Retrieved from Badan Pusat Statistik: https://www.bps.go.id/indikator/indikator/view_data_pub/0000/api_pub/YW40a21pdTU1cnJxOGt6dm43ZEdoZz09/da_03/1

Breiman, L. (1996). Bagging Predictors . *Machine Learning, 24*, 123-140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32. https://doi.org/10.1023/A:1010933404324

Buhl, H. U., & Heinrich, B. (2008). Valuing Customer Portfolios under Risk-Return-Aspects: A Model-based Approach and its Application in the Financial Services Industry. *Academy of Marketing Science Review Vol.12.* https://doi.org/10.5283/epub.23202

Chan, T. Y., Wu, C., & Xie, Y. (2011). Measuring the Lifetime Value of Customers Acquired from Google Search Advertising. *Marketing Science, 30*(5), 837-850. https://doi.org/10.1287/mksc.1110.0658

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 785-794). San Francisco. https://doi.org/10.1145/2939672.2939785

Cherchye, L., Rock, B. D., Dierynck, B., Kerstens, P. J., & Roodhoof, F. (2023). A DEA-based approach to customer value analysis. *European Journal of Operational Research.*

Chicco, D., Warrens, M., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science, 7*. https://doi.org/10.7717/peerj-cs.623

Cutler, A., Cutler, D., & Stevens, J. (2012). Random Forests. In C. Zhang, & Y. Ma, *Ensemble Machine Learning* (pp. 157-175). New York: Springer. https://doi.org/10.1007/978-1-4419-9326-7_5

Dahana, W., Miwa, Y., & Morisada, M. (2019). Linking lifestyle to customer lifetime value: an exploratory study in an online fashion retail market. *Journal of Business Research, Vol. 99*, (pp. 319-331). https://doi.org/10.1016/j.jbusres.2019.02.049

Daqar, M., & Smoudy, A. (2019). The Role Of Artificial Intelligence On Enhancing Customer Experience. *International Review of Management and Marketing*. https://doi.org/10.32479/IRMM.8166

Dhar, R., & Glazer, R. (2003). Hedging Customers. *Harvard Business Review, 81*(5), 86-92.

Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *arXiv*. https://doi.org/10.48550/arXiv.1702.08608

El Naqa, I., & Murphy, M. (2015). What Is Machine Learning? In I. El Naqa, & M. Murphy, *Machine Learning in Radiation Oncology* (pp. 3-11). Springer, Cham. https://doi.org/10.1007/978-3-319-18305-3_1

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

García, S., Luengo, J., & Herrera, F. (2014). *Data Preprocessing in Data Mining.* Springer. https://doi.org/10.1007/978-3-319-10247-4

Gelbrich, K., & Nakhaeizadeh, R. (2000). Value Miner: A Data Mining Environment for the Calculation of the Customer Lifetime Value with Application to the Automotive Industry. In R. L. Plaza, *Lecture Notes in Computer Science(), vol 1810* (pp. 154-161). Springer.

Glady, N., Lemmens, A., & Croux, C. (2015). Unveiling the relationship between the transaction timing, spending and dropout behavior of customers. *International Journal of Research in Marketing, 32(1).*, (pp. 78 - 93). https://doi.org/10.1016/j.ijresmar.2014.09.005

Gramegna, A., & Giudici, P. (2020). Why to Buy Insurance? An Explainable Artificial Intelligence Approach. *Risks, 8*(4), 137. https://doi.org/10.3390/risks8040137

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv., 51*(5), 1-42. https://doi.org/10.1145/3236009

Guo, R., Zhao, Z., Wang, T., Liu, G., Zhao, J., & Gao, D. (2020). Degradation State Recognition of Piston Pump Based on ICEEMDAN and XGBoost. *Applied Sciences, 10*(18). https://doi.org/10.3390/app10186593

Gupta, S., & Lehmann, D. (2003). Customers as Assets. *Journal of Interactive Marketing, 17*(1), 9-24.

Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research, 41*(1), 7 - 18. https://doi.org/10.1509/jmkr.41.1.7.25084

Haag, F., Hopf, K., Vasconcelos, P. M., & Staake, T. (2022). Augmented cross-selling through explainable AI -- a case from energy retailing. *Thirtieth European Conference on Information Systems (ECIS)*. https://doi.org/10.48550/arXiv.2208.11404

Hai-wei, W., Ming-hui, J., & Ya-lin, W. (2006). Adding Risk in Measuring Customer Value Using Bivariate Hierarchical Bayesian Approach. *International Conference on Management Science and Engineering.* Lille: IEEE. https://doi.org/10.1109/ICMSE.2006.313869

Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics, 106*(2), 787-804. https://doi.org/10.1007/s11192-015-1798-9

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning.* New York: Springer. https://doi.org/10.1007/978-0-387-84858-7

Hogan, J. E., Lehmann, D. R., Merino, M., Srivastava, R. K., Thomas, J. S., & Verhoef, P. C. (2002). Linking Customer Assets to Financial Performance. *Journal of Service Research, 5*(1), 26-38. https://doi.org/10.1177/1094670502005001004

Homburg, C., Steiner, V. V., & Totzek, D. (2009). Managing Dynamics in a Customer Portfolio. *Journal of Marketing Vol.73* (pp. 70-89). American Marketing Association. https://doi.org/10.1509/jmkg.73.5.70

Hwang, H., Jung, T., & Suh, E. (2004). An LTV Model and Customer Segmentation Based on Customer Value: a Case Study on the Wireless Telecommunication industry. *Expert Systems with Applications, 26*, 181-188.

IBM. (n.d.). *What is Random Forest?* Retrieved 06 29, 2023, from IBM Web site: https://www.ibm.com/topics/random-forest

Iglewicz, B., & Hoaglin, D. C. (1993). *How to Detect and Handle Outliers.* ASQC Quality Press.

Janzing, D., Minorics, L., & Blöbaum, P. (2019). Feature relevance quantification in explainable AI: A causal problem. *arXiv*. https://doi.org/10.48550/arXiv.1910.13413

Janzing, D., Minorics, L., & Blöbaum, P. (arXiv:1910.13413). Feature relevance quantification in explainable AI: A causal problem. *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS*. https://doi.org/0.48550/arXiv.1910.13413

Jensen, K. (2012, April 26). Modeler CRISP-DM. Retrieved from ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerCRISPDM.pdf

Jhaveri, S., Khedkar, I., Kantharia, Y., & Jaswal, S. (2019). Success Prediction using Random Forest, CatBoost, XGBoost and AdaBoost for Kickstarter Campaigns. *3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 1170-1173. https://doi.org/10.1109/ICCMC.2019.8819828

Juhl, H. J., & Christensen, M. (2013). Portfolio optimization and performance evaluation: An application to a customer portfolio. *Journal of Marketing Analytics, 1*, 156-173. https://doi.org/10.1057/jma.2013.11

Kang, J., Alejandro, T., & Groza, M. (2015). Customer–company identification and the effectiveness of loyalty programs. *Journal of Business Research, 68*, 464-471. https://doi.org/10.1016/J.JBUSRES.2014.06.002

Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Trans. Knowl. Discov. Data, 6*(4), 1-21. https://doi.org/10.1145/2382577.2382579

Kleinbaum, D. G., Dietz, K. G., Klein, M., & Klein, M. (2002). *Logistic regression.* Springer-Verlag. https://doi.org/10.1007/978-1-4419-1742-3

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Articial Intelligence , 4*(2), 1137-1145.

Kraus, S., Breier, M., & Lim, W. (2022). Literature reviews as independent studies: guidelines for academic practice. *Rev Manag Sci, 16*, 2577–2595. https://doi.org/10.1007/s11846-022-00588-8

Kumar, V. (2018). A Theory of Customer Valuation: Concepts, Metrics, Strategy,and Implementation. *Journal of Marketing, 82*, 1-19. https://doi.org/10.1509/jm.17.020

Kumar, V., & Reinartz, W. (2016). Creating Enduring Customer Value. *Journal of Marketing, 80*(6), 36-68. https://doi.org/10.1509/jm.15.0414

Kumar, V., Venkatesan, R., Bohling, T., & Beckmann, D. (2008). Practice prize Report—The power of CLV: Managing customer lifetime value at IBM. *Marketing Science, 27*(4), 585-599. https://doi.org/10.1287/mksc.1070.0319

Kundisch, D., Sackmann, S., & Ruch, M. (2008). Transferring Portfolio Selection Theory to Customer Portfolio Management – The Case of an e-Tailer. *:FinanceCom 2007, LNBIP 4* (pp. 32-49). Berlin: Springer-Verlag.

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy 2021, 23*(18). https://doi.org/10.3390/e23010018

Lipton, Z. C. (2018). The Mythos of Model Interpretability. *Commun. ACM, 61*(10), 36-43. https://doi.org/10.1145/3233231

Liu, J., Zhang, Y., Wang, X., Deng, Y., & Wu, X. (2021). Dynamic Pricing on E-commerce Platform with Deep Reinforcement Learning: A Field Experiment. *arXiv*. https://doi.org/10.48550/arXiv.1912.02572

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *31st Conference on Neural Information Processing Systems*, 4768–4777.

Machado, M. R., & Karray, S. (2022a). Integrating Customer Portfolio Theory and the Multiple Source of Risk Approaches to Model RAR. *IFAC PapersOnLine 55-16*, (pp. 356-363).

Machado, M. R., & Karray, S. (2022b). Applying hybrid machine learning algorithms to assess customer risk adjusted revenue in the financial industry. *Electronic Conference Research and Application 56.* https://doi.org/10.1016/j.elerap.2022.101202

Mahesh, B. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR), 9*(1), 381-386. https://doi.org/ 10.21275/ART20203995

Marín Díaz, G., Galán, J., & Carrasco, R. (2022). XAI for Churn Prediction in B2B Models: A Use Case in an Enterprise Software Company. *Mathematics, 10*(20). https://doi.org/10.3390/math10203896

Méndez-Suárez, M., & Crespo-Tejero, N. (2021). Why do banks retain unprofitable customers? A customer lifetime value real options approach. *Journal of Business Research, 122*, 621-626. https://doi.org/10.1016/j.jbusres.2020.10.008

Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*.

Myttenaere, A. d., Golden, B., Grand, B. L., & Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing, 192*, 38-48. https://doi.org/10.1016/j.neucom.2015.12.114.

Norouzi, A., & Albadvi, A. (2016). A hybrid model for customer portfolio analysis in retailing. *Management Research Review Vol.39, No.6* (pp. 63-654). Emerald Group Publishing Limited. https://doi.org/10.1108/MRR-04-2014-0082

Oshiro, T., Perez, P., & Baranauskas, J. (2012). How Many Trees in a Random Forest? *Machine Learning and Data Mining in Pattern Recognition.* Berlin: Springer. https://doi.org/10.1007/978-3-642-31537-4_13

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45-77. https://doi.org/10.2753/MIS0742-1222240302

Petersen, J. A., & Kumar, V. (2015). Perceived Risk, Product Returns, and Optimal Resource Allocation: Evidence from a Field Experiment. *Journal of Marketing Research, 52*(2), 268-285. https://doi.org/10.1509/jmr.14.0174

Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv:2010.16061*, 37-63. https://doi.org/10.48550/arXiv.2010.16061

Pranckute, R. (2021). Web of ˙Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications, 9*(12). https://doi.org/10.3390/publications9010012

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *32nd Conference on Neural Information Processing Systems.* Montréal, Canada.

Purnomo, M. R., Azzam, A., & Khasanah, A. U. (2020). Effective Marketing Strategy Determination Based on Customers Clustering Using Machine Learning Technique. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1471/1/012023

Qi, J., Zhou, Y., Chen, W., & Qu, Q. (2012). Are customer satisfaction and customer loyalty drivers of customer lifetime value in mobile data services: a comparative cross-country study. *Information Technology and Management, Vol. 13 No. 4*, (pp. 281-296).

Reinartz, W., & Kumar, V. (2002). The mismanagement of customer loyalty. *Harvard Business Review, 80*(7), 86-95.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv:1602.04938*, 1135-1144. https://doi.org/10.48550/arXiv.1602.04938

Ruch, M., & Sackmann, S. (2009). Customer-Specific Transaction Risk Management in E-Commerce. *AMCIS 2009: Value Creation in E-Business Management pp 68–79. 36.* Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-03132-8_6

Ruch, M., & Sackmann, S. (2012). Integrating management of customer value and risk in e-commerce. *Inf Syst E-Bus Manage , 10*, 101-116. https://doi.org/10.1007/s10257-010-0152-2

Rust, R. T., Kumar, V., & Venkatesan, R. (2011). Will the frog change into a prince? predicting future customer profitability. *International Journal of Research in Marketing, 28*(4), 281-294. https://doi.org/10.1016/j.ijresmar.2011.05.003

Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing, 68*(1), 109-127. https://doi.org/10.1509/jmkg.68.1.109.24030

Ryals, L. (2003). Making customers pay: measuring and managing customer risk and returns. *Journal of Strategic Marketing*, (pp. 165-175). https://doi.org/10.1080/0965254032000133476

Ryals, L. J., & Knox, S. (2005). Measuring risk-adjusted customer lifetime value and its impact on relationship marketing strategies and shareholder value. *European Journal of Marketing Vol.39* (pp. 456-472). Emerald Group Publishing Limited. https://doi.org/10.1108/03090560510590665

Ryals, L., & Knox, S. (2007). Measuring and managing customer relationship risk in business markets. *Industrial Marketing Management, 36*, 823-833. https://doi.org/10.1016/j.indmarman.2006.06.017

Sackmann, S., Kundisch, D., & Ruch, M. (2010). Customer portfolio management in e-commerce: an analytical model for optimization,. *Management Research Review, 33*(6), 617-634. https://doi.org/ 10.1108/01409171011050226

Samek, W., & Müller, K.-R. (2019). Towards Explainable Artificial Intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller, *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (pp. 5-22). Springer.

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 526-534. https://doi.org/10.1016/j.procs.2021.01.19

Sefara, T. J. (2019). The Effects of Normalisation Methods on Speech Emotion Recognition. *International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, 1-8. https://doi.org/10.1109/IMITEC45504.2019.9015895

Sharma, A. (2020, May 12). *Random Forest vs Decision Tree | Which Is Right for You?* Retrieved from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/

Sharpe, W. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance, 19*(13), 425-42.

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 1310-1315.

Singh, S., & Singh, S. (2016). Accounting for risk in the traditional RFM approach. *Management Research Review, 39*(2), 215-234. https://doi.org/10.1108/MRR-11-2015-0272

Singh, S., Murthi, B. S., & Steffes, E. (2013). Developing a measure of risk adjusted revenue (RAR) in credit cards market: Implication for customer relationship management. *European Journal of Operational Research* (pp. 425-434). Elsevier B.V. https://doi.org/10.1016/j.ejor.2012.08.007

So, C. (2020). What Emotions Make One or Five Stars ? Understanding Ratings of Online Product Reviews by Sentiment Analysis and XAI. *Artificial Intelligence, 1st International Conference on Artificial Intelligence in HCI,.* Kopenhagen, Denmark: Springer.

So, M. C., Thomas, L. C., Seow, H.-V., & Mues, C. (2014). Using a transactor/revolver scorecard to make credit and pricing decisions. *Decision Support Systems, 59*, 143-151. https://doi.org/10.1016/j.dss.2013.11.002

Souza, J., & Leung, C. K. (2021). Explainable Artificial Intelligence for Predictive Analytics on Customer Turnover: A User-Friendly Interface for Non-expert Users. In M. Sayed-Mouchaweh, *Explainable AI Within the Digital Transformation and Cyber Physical Systems* (pp. 47-67). Springer, Cham. https://doi.org/10.1007/978-3-030-76409-8_4

Tan, X. S., . Yang, Z., Benlimane, Y., & Liu, E. (2020). Using Classification with K-means Clustering to Investigate Transaction Anomaly. *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 171-174. https://doi.org/10.1109/IEEM45057.2020.9309909

Tarasi, C. O., Bolton, R. N., Hutt, M. D., & Walker, B. A. (2011). Balancing Risk and Return in a Customer Portfolio. *Journal of Marketing Vol.75* (pp. 1-17). American Marketing Association.

Telkomsel. (2023, 6 21). *Annual Report.* Retrieved from Telkomsel: https://www.telkomsel.com/en/about-us/investor-relations

Tsai, C., Hu, Y., Hung, C., & Hsu, Y. (2013). A comparative study of hybrid machine learning techniques for customer lifetime value prediction. *Kybernetes, 42*(3), 357-370. https://doi.org/10.1108/03684921311323626

Van Lent, M., Fisher, W. C., & Mancuso, M. (2004). An Explainable Artificial Intelligence System for Small-unit Tactical Behavior. *In Proc. 16th Conf.*, 900-907.

Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing, 68*(4), 106-125. https://doi.org/10.1509/jmkg.68.4.106.42728

Viviani, J.-L., Komura, A., & Suzuki, K. (2021). Integrating dynamic segmentation and portfolio theories for better customer portfolio performance. *Journal of Strategic Marketing*. Routledge. https://doi.org/10.1080/0965254X.2021.1881148

Wangenheim, F. V., & Lentz, P. (2005). Customer Portfoilio Analysis : Applying Financial Risk and Volatility Measures to Customer Segmentation and Risk-Adjusted Lifetime Value Determination. https://doi.org/10.2139/ssrn.782064

Weld, D., & Bansal, G. (2019). The Challenge of Crafting Intelligible Intelligence. *Commun. ACM, 62*(6), 70-79. https://doi.org/10.48550/arXiv.1803.04263

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*(1), 79-82.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., & Zhao, D. (2019). Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. *8th CCF International Conference* (pp. 563-574). Dunhuang, China: Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-32236-6_51

Yılmaz Benk, G., Badur, B., & Mardikyan, S. (2022). A New 360° Framework to Predict Customer Lifetime Value for Multi-Category E-Commerce Companies Using a Multi-Output Deep Neural Network and Explainable Artificial Intelligence. *Information, 13*(8), 373. https://doi.org/10.3390/info13080373

Yun, C., & Yan, P. (2013). A customer portfolio model based on multi-phase marketing strategy and particle swarm optimization. *International Conference on Management*

Science and Engineering 20th Annual Conference Proceedings, (pp. 957-962). Harbin, China. https://doi.org/10.1109/ICMSE.2013.6586393.

Appendix 1

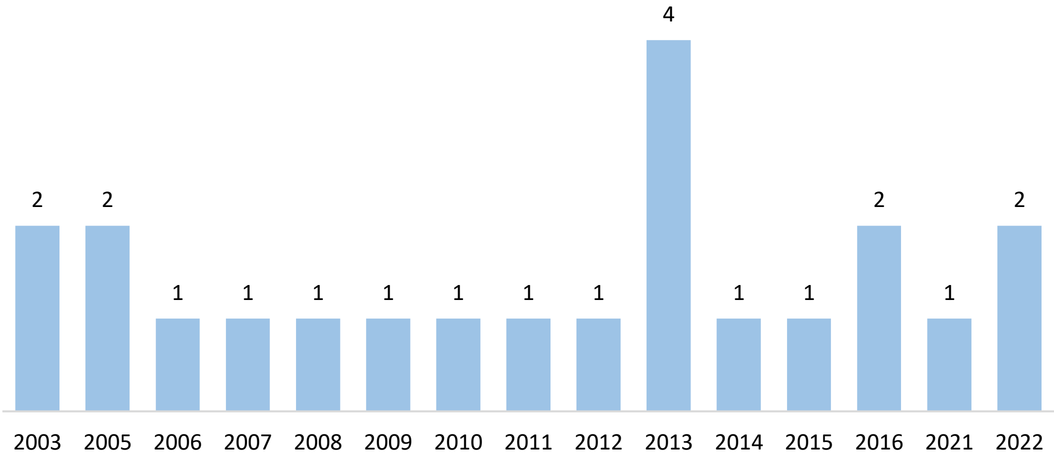


Figure 45 Detailed overview of publication per year

Appendix 2

Table 26 Sample data dictionary

Feature Name	Data Type	Description
msisdn	int	The mobile phone number associated with the customer.
poin	Int	The loyalty points earned by the customer.
total_revenue	Int	The total revenue generated by the customer.
total_recharge	Int	The total recharge conducted by the customer.
dominan_denomination	Int	The dominant or most frequent recharge denomination used by the customer.
Dominan_recharge_channel	obj	The dominant channel used by the customer to recharge.
avg_3m	Int	The average of last 3months revenue by the customer.
voice/sms/broadband/ digital_services/ir/others	Int	The type of services used by the customer
revenue_xxx	Int	The revenue generated by using the service 'xxx'
transaction_xxx	Int	The total number of transactions for the service 'xxx' made by the customer
duration_xxx	Int	The total duration of 'xxx' made by the customer
volume_of_xxx	Int	The total volume of 'xxx'
total_day_of_xxx	Int	The total number of days customer using 'xxx' service
xxx_package	Int	The package related for service purchased by customer
last_recharge/revenue	Datetime	The last recorded date of such transaction
xxx_std	Int	Indicate the standard deviation of such value over the observation period
xxx_pct_change	Int	Indicate the average change (in pct) of such value over the observation period
status	Obj	The status of the customer's account or subscription.
kabupaten	Obj	The name of the district or regency where the customer is located.
kelurahan	Obj	The name of the sub-district or neighborhood where the customer is located.
dominan_revenue	Obj	The dominant or most frequent revenue type or category for the customer.
gender	Obj	The gender of the customer.
Brand_x	Obj	The pricing plan/brand selected by customer.
region_lacci_xxx	Obj	The region or location identifier based on the LAC-CI (Location Area Code-Cell Identity).
los	int	The length of stay or tenure of the customer.
device_type	Int	The type or model of the customer's device.
Device_data_capable	Int	Indicates whether the customer's device is capable of using data services.
Device_os_priority	Int	The priority or preference given to the customer's operating system.
Device_network_prio	Int	The priority or preference given to the customer's network connection.
data_usage_category	Int	The category or classification of the customer based on their data usage.
Most_used_apps	Int	The name of the app ranked first by the customer.

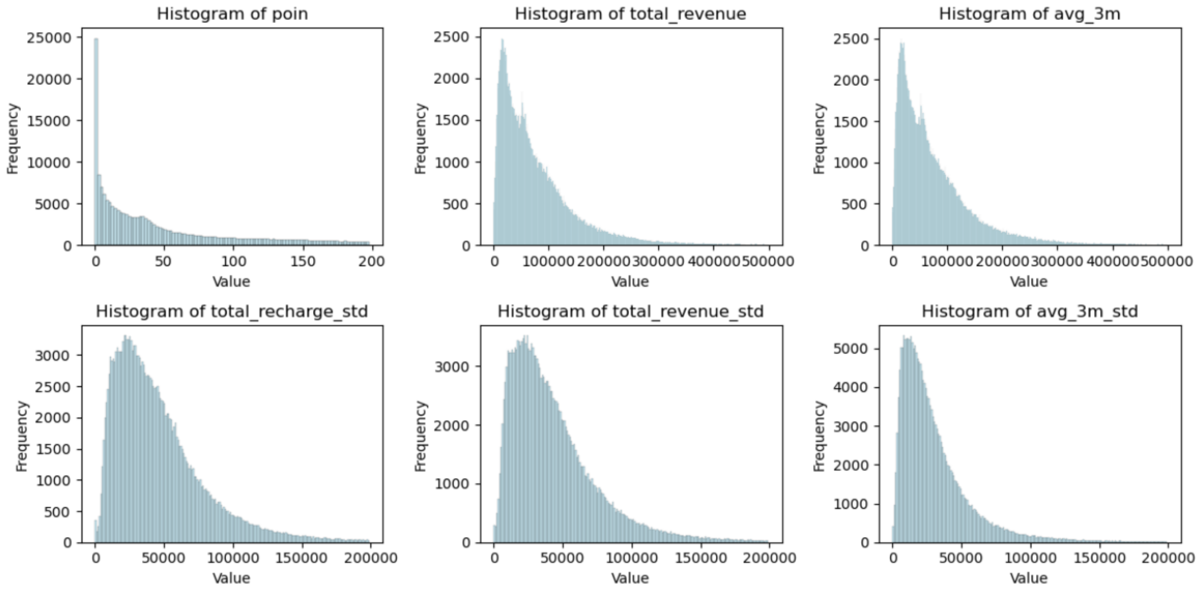


Figure 46 Histogram of Sample Features

Table 27 Sample Feature Correlation with Target Value

Feature name	RAR_total_revenue	abs_value	Rank
avg_3m	0.839	0.839	1
poin	0.729	0.729	2
total_revenue	0.719	0.719	3
avg_3m_std	0.684	0.684	4
total_revenue_std	0.650	0.650	5
total_recharge_std	0.612	0.612	6
revenue_broadband	0.596	0.596	7
revenue_broadband_std	0.541	0.541	8
revenue_digital_services	0.461	0.461	9
revenue_digital_services_std	0.442	0.442	10
volume_broadband	0.438	0.438	11
dominan_denomination	0.403	0.403	12
segment_tier_poin_A	0.387	0.387	13
customer_tier_A	0.368	0.368	14
day_of_digital_services	0.367	0.367	15
…	…	…	…
voice_pct_change	-0.03	0.03	141
region_lacci_02	-0.03	0.03	142
most_used_apps_eCommerce	0.03	0.03	143
transaction_voice_international_call	0.03	0.03	144
dominan_recharge_channel_Retailer	-0.03	0.03	145
voice_package_A_user	0.03	0.03	146
device_data_capable_UNIDENTIFIED	-0.02	0.02	147
transaction_voice_off_net	0.02	0.02	148
transaction_voice_package_B	-0.02	0.02	149
region_lacci_12	-0.02	0.02	150
dominan_revenue_Others	-0.02	0.02	151

most_used_app_category_SocialNetwork	0.02	0.02	152
duration_others	0.02	0.02	153
dominan_revenue_Digital Services	-0.02	0.02	154
region_tier	0.02	0.02	155