HYBRID DEEP NETWORKS BASED ON PERIOD-SHIFT COSINE ANNEALING FOR CUSTOMER RETENTION PREDICTION IN TELECOM INDUSTRY

JOHNSON OLANREWAJU VICTOR

UNIVERSITI SAINS MALAYSIA

2024

HYBRID DEEP NETWORKS BASED ON PERIOD-SHIFT COSINE ANNEALING FOR CUSTOMER RETENTION PREDICTION IN TELECOM INDUSTRY

by

JOHNSON OLANREWAJU VICTOR

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

August 2024

ACKNOWLEDGEMENT

I express my profound gratitude, beginning with the Creator of the universe, *God*, whose infinite love and mercy have guided me thus far. I am humbled by the countless individuals through whom God worked, providing unwavering support throughout my life journey. To my late grandmother, Madam Eunice Awonusi, your role transcends words. You were truly God-sent, fighting persistently until your grandson saw the light of day. My late parents, Mr. Johnson Jimoh (*Easyway Photos*) and Mrs. A. Margaret Johnson, I thank you for your unwavering support. The dreams you envisioned for me materialized despite the challenges. I wish I could witness the joy in your faces. A special acknowledgment to my supportive siblings; I am deeply indebted for your significant contributions. Johnson Olugbenga, your fatherly role is commendable, and I appreciate your remarkable support. To my loving wife-Olabisi and children (*The 3 Mos*), I am profoundly grateful for your understanding, love, prayers, and unwavering support. Your sacrifice has given meaning to this journey. I always hear the question, "Daddy, when are you coming home?" Thank you for the sacrifices made. The entirety of my Ph.D. journey would lack vibrancy without the guidance of my incredible supervisor, Ts. Dr. Chew XinYing. Your unparalleled support and readiness to assist at all times are deeply appreciated. *I am grateful for your belief in me*. I greatly acknowledge Nigeria Government for providing the funding for this degree through TEFUND, Nigeria. Thanks are due to the Dean and all staff of the School of Computer Sciences at USM. I extend my gratitude to Prof E.A Fasakin, Prof O. Adetunmbi, Rev. S Ogunrinde, Mrs F. Ogunbodede, Dr. Ayokunle Ige, Mutiu Ganiyu, my lab partner Ige O. P, Obaseki Samuel, and many others too numerous to mention. Thank you, as we collectively achieved this significant milestone together!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALORITHMS

Page

LIST OF ABBREVIATIONS

1D-CNN	One-Dimensional Convolutional Neural Network
ADAB	Ada Boosting
ADAYSN	Adaptive Synthetic Sampling
AE	Autoencoder
AF	Activation Function
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AUC	Area Under ROC
BCE	Binary Cross-Entropy
CALR	Cosine Annealing LR
CCE	Categorical Cross-Entropy
CDR	Call Details Records
CF	Cost Function
CHAID	Chi-square Automatic Interaction Detection
CIRW	Class-Imbalance-Ratio-Weight
CLV	Customer Lifetime Value
CNN	Convolutional Neural Network
CoRCAT	Cost-enabled Residual Convolutional-Attention Tree-Structured Multilayer
CRM	Customer Relationship Management
CRP	Customer Retention Prediction
CSL	Cost-Sensitive Learning
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Trees
EDA	Exploratory Data Analysis
FCL	Fully Connected Layers
GB	Gradient Boosting
GBDT	Gradient Boosting Decision Trees
Gmean	Geometric Mean.
GPU	Graphical Processing Unit

GSMOTE	Geometric Synthetic Minority Over-sampling Technique
ICOTE	Immune centroids oversampling technique
IoE	Internet of Expression
IoTs	Internet of Things
IQR	interquartile rule
KNN	K-Nearest Neighbour
LF	Loss Function
LGB	Light Gradient Boosting
LIME	Local Interpretable Model-agnostic Explanation
LR	Learning Rate
LR	Logistics Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multilayer Perceptron
MTDF	Mega-Trend-Diffusion Function
MWMOTE	Majority Weighted Minority Oversampling Technique
NB	Naïve Bayes
NLP	Natural Language Processing
NN	Neural Network
NP	Nondeterministic Polynomial
PCA	Principal Component Analysis
PPV	Positive Predictive Value
PR	Precision-Recall
ps-CALR	Periodic Shift Cosine Annealing LR
RCA	Residual Convolutional-Attention
RELU	Rectifier Linear Unit
RF	Random Forest
RMSProp	Root Mean Square Propagation
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristics
ROS	Random Oversampling
RUS	Random Undersampling
SAE	Stacked Autoencoders
SGD	Stochastic Gradient Descent

SHAP	Shapley Additive explanations
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
t-SNE	t-distributed Stochastic Neighbor Embedding t-SNE
TNR	True Negative Rate
TPR	True Positive Rate
TSM	Tree-Structured Multilayer
XGB	eXtreme Gradient Boosting

LIST OF APPENDICES

RANGKAIAN DALAMAN HIBRID BERDASARKAN TEMPOH-SYIF PENYEPUHLINDAPAN KOSINUS UNTUK RAMALAN PENGEKALAN PELANGGAN DALAM INDUSTRI TELEKOMUNIKASI

ABSTRAK

Dalam landskap dinamik Ramalan Pengekalan Pelanggan (RPP), keperluan untuk mengarahkan usaha pemasaran dan promosi secara strategik ke arah pelanggan sasaran tidak pernah menjadi lebih penting. Mengenal pasti penunjuk yang berpotensi dan meneroka kaedah pengekalan inovatif secara berterusan menjadi yang paling penting. Walau bagaimanapun, cabaran utama adalah pelanggan yang menamatkan perkhidmatan mereka jarang dikenali di kalangan mereka yang setia yang membawa kepada masalah ketidakseimbangan. Pembelajaran Mesin Konvensional (MK), dengan pergantungan lazimnya pada pengekstrakan ciri dan kaedah pensampelan data, termasuk teknik sensitif kos, bergelut dengan isu seperti pemasangan berlebihan, kerumitan pengiraan dan penekanan yang tidak wajar pada kes yang jarang berlaku. Teknik Pembelajaran Dalam (PM) yang digunakan pada RPP menjanjikan pengekstrakan ciri automatik berbanding kaedah buatan tangan yang digunakan dalam MK. Walau bagaimanapun, sifat tidak sensitif kos, Kadar Pembelajaran (KP) yang dipilih dengan sewajarnya untuk penumpuan yang lebih baik, dan pembelajaran ciri berkualiti dalam PM masih menimbulkan cabaran. Tesis ini memperkenalkan Berat Nisbah Ketidakseimbangan Kelas (BNKK) yang direka untuk menangani masalah ketidakseimbangan dalam pengelas PM tanpa menanggung kos pengiraan tambahan atau kehilangan simetri data. Selain itu, ia mencadangkan kaedah Tempoh-Syif Kosinus Penyepuhlindapan KP (ts-KPKP) untuk menangani dinamik KP semasa latihan PM, dengan itu meningkatkan generalisasi. Akhir sekali, PM hibrid, menggabungkan saraf berbilang lapisan yang dipertingkatkan dan rangkaian saraf konvolusional satu dimensi, dibangunkan untuk mempelajari ciri yang dipertingkatkan untuk analisis pengekalan pelanggan. hibrid ini disepadukan dengan kedua-dua BNKK dan ts-KPKP. Percubaan pada data yang tersedia secara terbuka menunjukkan bahawa model yang dicadangkan mencapai skor F1 sebanyak 74.58% pada data IBM, 91.93% pada data Iran dan 88.80% pada data Jingga, garis dasar yang lebih baik, kaedah MK konvensional dan model dalam sedia ada. pengajian. RPP hibrid mempersembahkan prestasi terkini dan memberikan pandangan praktikal untuk mengoptimumkan strategi pengekalan pelanggan dalam landskap industri telekomunikasi yang berkembang.

HYBRID DEEP NETWORKS BASED ON PERIOD-SHIFT COSINE ANNEALING FOR CUSTOMER RETENTION PREDICTION IN TELECOM INDUSTRY

ABSTRACT

In the dynamic landscape of Customer Retention Prediction (CRP), the imperative to strategically direct marketing and promotion efforts towards targeted customers has never been more crucial. Identifying potential churn indicators and continually exploring innovative retention methods becomes paramount. However, a major challenge is customers terminating their services are rarely known among the loyal ones leading to an imbalance problem. Conventional Machine Learning (ML), with its prevalent reliance on feature extraction and data sampling methods, including cost-sensitive techniques, grapples with issues such as overfitting, computational complexity, and an undue emphasis on rare cases. Deep Learning (DL) techniques applied to CRP is promising for automatic feature extraction compared to the handcrafted method used in ML. However, non-cost-sensitive nature, appropriately chosen Learning Rate (LR) for better convergence, and quality feature learning in DL models still pose challenges. This thesis introduces a Class Imbalance Ratio Weight (CIRW) designed to tackle the imbalance problem in DL classifiers without incurring additional computational costs or loss of data symmetry. Additionally, it proposes a novel Period-Shift Cosine Annealing Learning Rate (ps-CALR) method to address LR dynamics during DL model training, thereby enhancing generalization. Finally, a hybrid DL model, combining an improved multilayer perceptron and a one-dimensional convolutional neural network, is developed to learn improved features for customer retention analysis. This hybrid model is integrated with both the CIRW and ps-CALR. Experiments on publicly available datasets show that the proposed model achieved an F1-score of 74.58% on the IBM dataset, 91.93% on the Iranian dataset, and 88.80% on the Orange dataset, outperforming baseline models, conventional ML methods, and models in existing studies. The hybrid CRP model presents state-of-the-art performance and provides practical insights into optimizing customer retention strategies in the evolving landscape of the telecommunications industry.

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter provides an overview of the research. Section 1.2 presents the background of the study, while Section 1.3 discusses the motivation behind the research. The problem statement is outlined in Section 1.4, followed by the research questions in Section 1.5. Section 1.6 focuses on the aim and objectives of the research, and Section 1.7 elaborates on the scope of the study. The expected contributions to knowledge are highlighted in Section 1.8. Finally, Section 1.9 presents the overall outlines research.

1.2 Background of the Study

The digital transformation witnessed worldwide in the last decade compared to other decades has grown exponentially. A constant and massive deployment of technological innovations is helping to rethink how humans view their activities and the environment. The upsurge in computing innovative technology, including cloud computing, Internet of Things (IoTs), Internet of Expression (IoE), open-source software, and Blockchain technology, to mention but a few, have paved the way and pushed virtually every existing human activity to “On-Demand” services. Inadvertently, telecommunication technology will continue to play a pivotal role in ensuring seamless delivery of the “On-Demand” services and fostering how people connect or socialize. The majority of people in the world connect through a telecommunication operator. For instance, Global Monitor reported that out of 31.83 million Malaysians, 25.08 million are Internet users as of 2018. The growth is expected to push e-commerce to above 20.8% in 2020 and beyond (Global Monitor, 2021).

The telecom market is projected at RM38.81 billion from 2023 to RM42.27 billion in 2028. At the global level, there are around 5.44 billion Internet users, amounting to 67.1% of the global population (Petrosyan, 2024). The increasing Internet penetration is a direct consequence of mobile telecom today. Several players in the telecommunications industry, both at the national and international levels, act as service providers for the worldwide population. For example, the key players in Malaysia include Maxis, Digi, Celcom, Edotco and Sacofa Sdn Bhd. The benefit from the above projections is the vast market opportunity available for telecom operators and indirectly for other companies rendering consumer needs as services.

Moreover, customers leaving one service provider for another is a primary concern among the frequent and significant dangers facing telecom companies as service providers. According to (Rohit, 2024), the telecommunications industry has a customer retention rate of 78%, which ranks among the top sectors like media (84%) and professional services (84%), as presented in Table 1.1, underscoring the urgency for customer retention strategies. Also, the research by Ribeiro et al. (2023) highlights a steady increase in published papers on customer churn, especially accelerating from 2015 onwards, driven by advancements in data analytics and the growing importance of customer retention, as shown in Figure 1.1. The indication is the potential of research interest and existence of research gaps, suggesting opportunities for researchers to delve into new dimensions such as emerging customer segments and the role of Artificial Intelligence (AI).

Meanwhile, the implications of customer churn are profound, often resulting from overburdened infrastructure, network failures, inadequate customer service, and software issues. These factors contribute to customer dissatisfaction and subsequent churn. Change in customer experience owing to digital technology dynamics and stiff competition from other service providers (Spanoudes & Nguyen, 2017) are other factors leading to customer churn.

Table 1.1 Customer retention rates across different companies (Rohit, 2024)

Industry	Customer Retention Rates (%)
Media	84
Professional Services	84
Automotive & Transportation	83
Insurance	83
IT Services	81
Construction & Engineering	80
Financial Services	78
Telecommunications	78
Healthcare	77
IT & Software	77
Banking	75
Consumer Services	67
Manufacturing	67
Retail	63
Hospitality, Trave, Restaurants	55

The consequences of customer churn include direct revenue losses, and the substantial costs associated with acquiring new customers. Therefore, telecom companies must prioritize improving customer experience, leveraging advanced predictive models for customer retention, and addressing the root causes of churn. By doing so, they can significantly enhance their retention rates, secure their revenue streams, ensure their long-term survival in a highly competitive market, and organic growth through word-of-mouth referrals. So, when a customer leaves, there are one million and one questions on why and how the customer left because companies are aware that retaining existing customers imparts their survival.

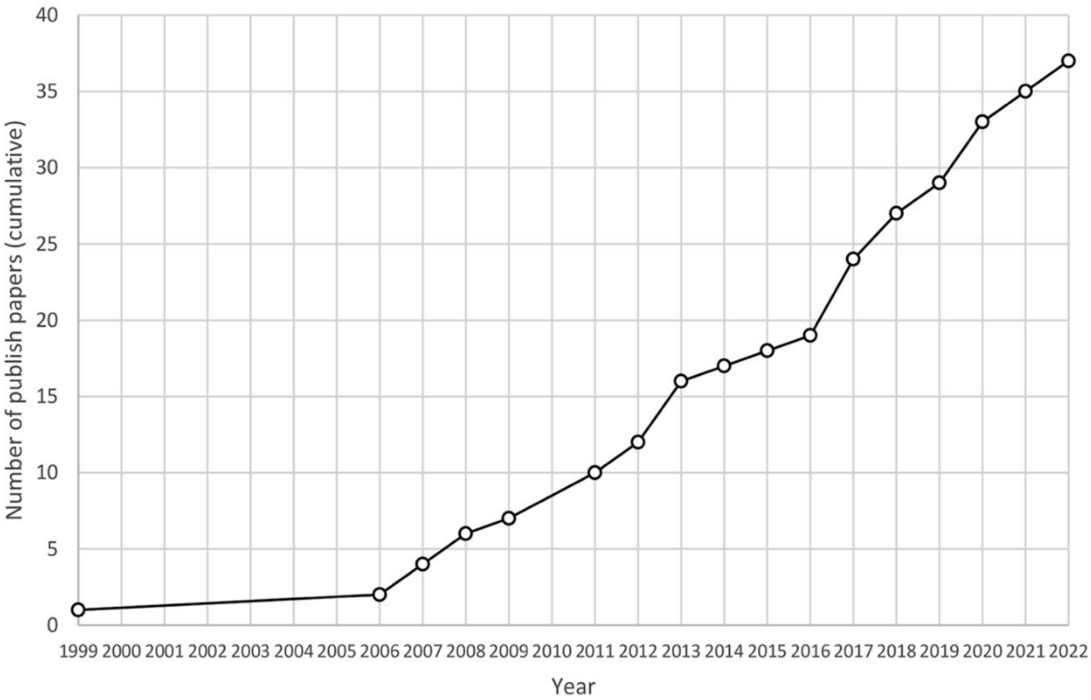

Figure 1.1 Trend of customer retention prediction research (Ribeiro et al., 2023)

It is, therefore, essential for the company to have prior knowledge of its customers to refine its marketing strategy. Additionally, implementing a system that can predict when customers are likely to switch network providers is crucial. By adopting these strategies, the company can reduce the risk of customer attrition and gather valuable business intelligence. Notably, Machine Learning (ML) approaches have shown promising results in enhancing customer retention efforts (Maw et al., 2019; Nhu et al., 2022).

Meanwhile, customers eager to leave for another service provider are rarely few among the loyal ones, making accurate prediction difficult because the ML algorithm tends to bias the majority class over the minority, as shown in Figure 1.2. This is known as a data imbalanced problem. Existing approaches to solving data imbalance are overwhelmed with drawbacks, including overfitting, loss of information, high computing cost, non-data-centricity, and overemphasis placed on rare cases (Kaur et al., 2019; Nguyen & Duong, 2021).

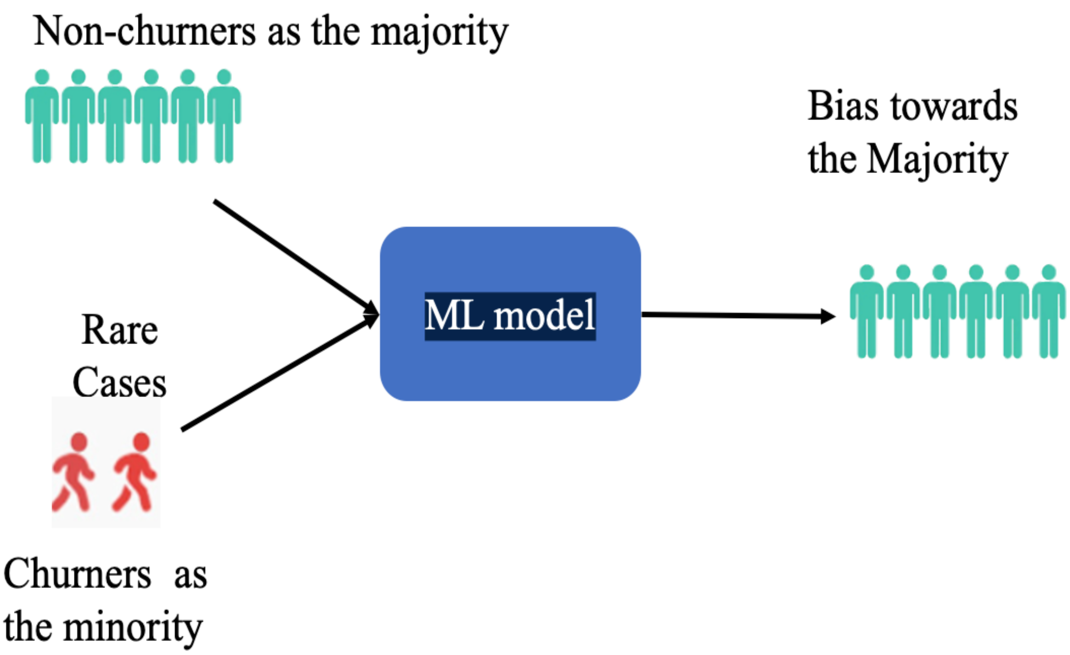

Figure 1.2 ML behaviour towards imbalance customer data during training

Another concern is that existing ML methods depend on feature engineering techniques to provide relevant features for training the customer data to obtain the best possible performance. The over-dependence of ML on feature engineering is computationally expensive, time-consuming, laborious, and often requires domain experts despite helping to reduce overfitting in ML models (Li et al., 2022; Umayaparvathi & Iyakutti, 2017). Nevertheless, no feature engineering could be considered the best. When applied to imbalance data, the feature engineering task is a Nondeterministic Polynomial (NP) problem. Over time, Deep Learning (DL) has gained wide attention for its inherent capability of performing automatic feature extraction vital for analysis of the input data. This mechanism makes DL sufficiently handy to train non-image structured data without requiring a prior feature engineering process.

Moreover, the DL model is subject to being trapped in local minima during the optimisation process, as shown in Figure 1.3. The ideal is the DL model escaping the local minima to a global minimum, where the optimal solution lies. Carefully choosing the Learning Rate (LR) can help solve this problem (Jepkoech et al., 2021; Wu & Liu, 2023). Hence, many LR scheduling techniques used in deep networks are flexible enough to balance rapid initial progress with stable convergence during training. Therefore, a high LR helps DL models achieve strong generalization but results in slow model training. Conversely, a lower LR helps the model, which may occasionally get stuck in local minima, to search for a global minimum, though it often leads to poor generalization to the training data. Existing Cosine Annealing LR (CALR), among others, is known for starting with an initial high LR and oscillating to a very small LR (Huang et al., 2017; Loshchilov & Hutter, 2017). This approach fulfils faster convergence of the model to the global minimum; however, it lacks a complete exploration of the loss landscape. Therefore, it may not provide a better model generalisation.

Furthermore, conventional ML's dependence on feature engineering is bypassed with the unsupervised feature engineering inherent in the DL model, as previously mentioned. Consequently, the DL model may experience a drawback in performance if its feature extraction process is not effective and robust enough to help in enhancing performance. Recent DL model advancements advocate mechanisms to enhance its selective focus on important features to improve training efficiency and prediction accuracy (Devassy & Antony, 2023; Niu et al., 2021). However, these concepts are not widely used in the literature on the DL telecom customer retention model.

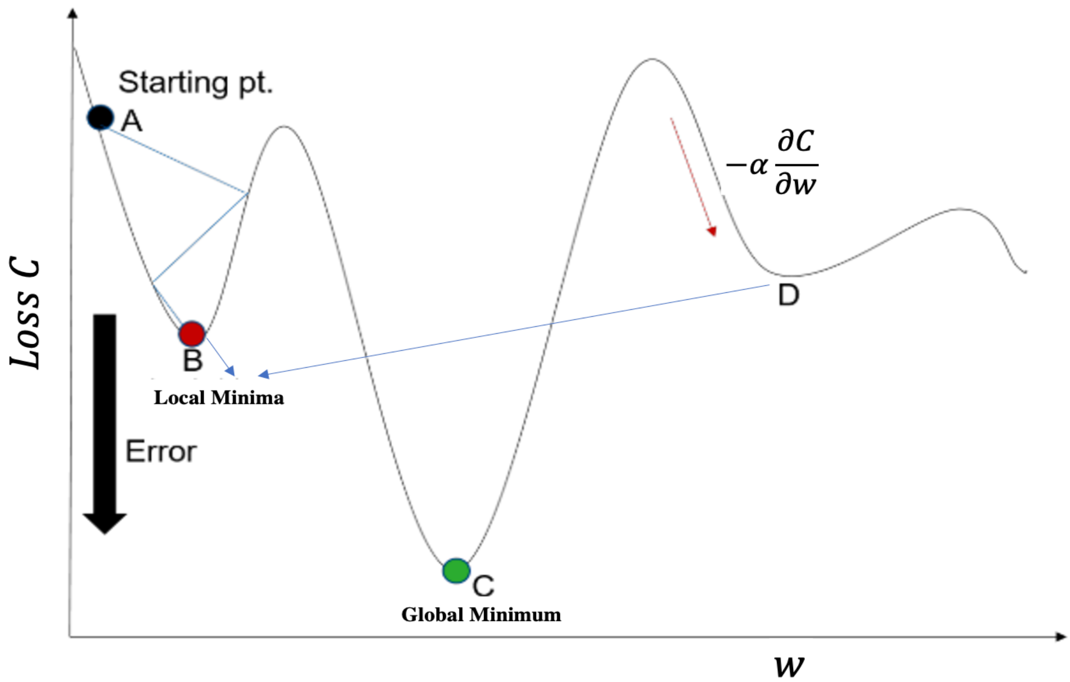

Figure 1.3 Gradient descent process with different minima points adapted from (Nie et al., 2022)

An ideal model for Customer Retention Prediction (CRP) in the telecom industry should be cost-effective, capable of handling imbalanced datasets, and adept at overcoming local minima challenges through LR scheduling. Moreover, it requires robust feature engineering to enhance predictive accuracy. To address these requirements, this research proposes a novel hybrid DL model named Cost-enabled Residual Convolutional-Attention Tree-Structured Multilayer (CoRCAT).

CoRCAT integrates a Residual-Convolutional-Attention (RCA) block with a Tree-Structured Multilayer (TSM) architecture. By incorporating a cost-metric, the model effectively addresses data imbalance, while learning rate scheduling accelerates the training process. This innovative framework represents a significant advancement in telecom CRP modelling, surpassing existing state-of-the-art approaches.

1.3 Research Motivation

Telecommunication companies face the persistent challenge of customer churn, where customers switch from one service provider to another, driven by factors such as network failures, inadequate customer service, and competitive pressures. The significance of customer retention is crucial for safeguarding revenue streams, reducing high costs associated with acquiring new customers, enhancing company's long-term viability in the face of competition, and leading to organic growth through word-of-mouth referrals. Meanwhile, CRP model is hampered with ability to accurately identify the fewer customers intending to leaver to another service provider.

Recent CRP studies have utilized various methods to address imbalanced data, including data-level methods (Fujo et al., 2022; Hartati et al., 2018), algorithm-level approaches (Tuck et al., 2020; Wong et al., 2020), cost-sensitive techniques (Nguyen & Duong, 2021), and hybrid strategies (Salunkhe & Mali, 2018). CRP models have also evolved from conventional single models (Jain et al., 2020) to ensemble or hybrid models combining ML and DL (Li et al., 2022; Saghir et al., 2019; Wangperawong et al., 2016). The use of DL in recent times still relies on existing approaches to address data imbalance, which suffer from overfitting, the introduction of synthetic data, and high computational costs. A cost-effective approach that works inherently with DL models without disrupting data symmetry or incurring additional computational costs has not yet been explored for CRP in the literature.

Also, the reliance on feature engineering hinders the performance of ML models. DL models, which can automatically extract features (Li et al., 2022), offer a significant advantage by reducing the need for manual feature engineering. However, DL models face their own challenges, such as optimization issues like local minima during training, which can negatively impact their generalization capabilities (Liu, 2022). This contrast highlights that while DL models alleviate some of the burdens associated with ML models, they introduce new complexities that must be addressed to fully harness their potential in tasks like customer retention prediction. Existing Learning Rate (LR) scheduling techniques such as Cosine Annealing LR (CALR) (Loshchilov & Hutter, 2017) offer some solutions but still fall short in providing a complete exploration of the loss landscape, which is necessary for better model generalization. Hence CALR can be improved to achieve faster convergence and better generalization.

Additionally, the over-dependence on traditional feature engineering methods in ML mentioned earlier is not only time-consuming and laborious but also requires domain expertise, which is impractical for large-scale applications (Li et al., 2022). A DL model that can effectively perform feature engineering to improve customer CRP in the telecom industry is highly desirable. However, methods for enhancing feature learning by capturing both global and local features during deep training have not been thoroughly explored in the literature.

Motivated by these challenges, this research seeks to develop a cost-enabled hybrid model that addresses the imbalanced data problem, incorporates an improved CALR scheduling technique to overcome local minima issues, and ultimately enhancing the automatic feature extraction capabilities to improve performance.

1.4 Problem Statement

CRP is primarily a binary classification task. A binary classification is a specific type of supervised learning that categorises input data into one of two classes or categories denoted as the target label in the dataset under study. The class label is task-specific, and in the case of customer retention in telecom, it is denoted as churners and non-churners, regardless of the ontology definition used in the data description. It is a well-known fact that the class distribution is generally uneven, i.e., the number of customers in each class varies dramatically, resulting in majority and minority instances, principally addressed as a data imbalanced problem.

Data imbalanced problems, common across business, healthcare, and Natural Language Processing (NLP), are pervasive in ML. It affects the accurate performance of ML methods because prediction is biased towards the majority class, whereas it is the minority class that forms the focal interest. The implications can be devastating in most cases and life-threatening. For instance, predicting a sick patient as non-sick, non-credit worthy as worthy, or churners as non-churners. These circumstances are highly misleading.

Existing data sampling techniques include undersampling the majority class or oversampling the minority class(es). The undersampling method, such as NearMiss (Mqadi et al., 2021), has a challenge of information loss due to many instances of the majority class removed that can contribute meaningfully during training. In the case of oversampling, such as the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) and Geometric Synthetic Minority Oversampling Technique (GSMOTE) (Douzas & Bacao, 2019), the ML model is prone to overfitting because minority class(es) instances are augmented with synthetic samples. Recent studies have demonstrated a promising improvement in synthetic data generation for solving ML problems (Frid-Adar et al., 2018; S. Jain et al., 2022). Hybrid data sampling methods are further proposed in the literature (Kaur & Gosain, 2020; Ye et al., 2021) but are liable to additional overhead costs and data symmetry loss. The use of Support Vector Machine (SVM) and skewed-insensitive Bayesian models are faced with computational complexity over the training of order $O(n^2)$ and $O(n^3)$, where n

represents the input size in terms of number of data points (Cervantes et al., 2020). In addition, reducing variance and eliminating bias in an imbalance data was introduced through ensemble (Hartati et al., 2018; Lal & Kumar, 2021; Shumaly et al., 2020). The effectiveness of these variants techniques varies and, consequently, is non-data centric.

Surprisingly, cost-sensitive techniques are less used compared to other sampling techniques. The drawback is due to difficulty in identifying and assigning the cost of misclassification. Nevertheless, it is a better approach to data or algorithmic-based techniques because it falls between them. The cost-sensitive classifier favours the rare class by assigning more significant misclassification costs and attempting to reduce the overall cost errors of both classes during training. It does not hamper the original symmetry of the data. Existing approaches to cost-sensitive implementation, including focal loss and weighted cross entropy loss implemented in Nguyen et al. (2021) and cost matrix (Thakkar et al., 2022; Wong et al., 2020), are promising. However, they are characterised by high computing overhead, a chance of perturbing the gradient, and an overemphasis on rare class(es). Assigning weights in the weighted cross-entropy is crucial for handling imbalanced data in deep learning (DL) models. However, these weights are not always based on class balance. To address this issue, it is essential to propose a method that effectively manages imbalanced data without incurring additional computational costs. This approach can be seamlessly integrated into a DL model to enhance performance.

Moreover, DL has recently gained significant attention for solving a wide range of problems, including those involving non-image data (Pouyanfar et al., 2019; Sjarif et al., 2020). Both Multilayer Perceptron (MLP) and One-Dimensional Convolutional Neural Network (1D-CNN) are well-regarded for their effectiveness in classifying non-image data (Li et al., 2022; Sjarif et al., 2020). These DL techniques are particularly useful because they are less reliant on feature engineering, making them practical for a variety of applications. They possess the inherent capability for feature extraction (Li et al., 2022; Umayaparvathi & Iyakutti, 2017). Meanwhile, the learning mechanism of DL models as an optimisation problem depends on its Cost Function (CF) or Loss Function (LF) (Wu & Liu, 2023). The model minimises the LF repeatedly until it achieves convergence and generalisation over the data in the loss landscape. The LR, among DL hyperparameters, represents the small term in the iterative update of the cost function for achieving minimisation (Wu & Liu, 2023). However, one prevalent challenge with DL optimisation is getting stuck in the local minima, which is against the desired global minimum (Ruder, 2016). A carefully chosen LR, therefore, aids faster convergence and better generalisation of DL models.

The existing CALR method proposed by Loshchilov and Hutter (2017) helps manage LR oscillations during training by allowing the LR to fluctuate between high and low values, which can accelerate convergence. Liu (2022) enhanced this approach by introducing a linear warm-up period before the oscillations, starting with a very small LR and gradually increasing it to a higher value. This linear warm-up is believed to improve the model's generalization during training. However, despite the faster convergence achieved with these methods, they still fall short in providing faster convergence for DL models due to inadequate exploration and exploitation of the loss landscape around the global minimum. Developing approach to improve on the existing CARL is an area requiring sufficient contribution in the literature.

In addition to optimising DL convergence using the LR scheduling policy, it is crucial that the automatic feature extraction performed by the model ultimately enhances its performance. It is well known that feature engineering facilitates ML performance by removing irrelevant feature(s) in data (Ramos-Pérez et al., 2022). DL models can automatically extract relevant features from data, thereby removing the identified drawbacks of feature engineering (Li et al., 2022; Umayaparvathi & Iyakutti, 2017). For instance, Li et al. (2022) uses 1D-CNN to perform automatic feature extraction and Gradient Boosting Decision Trees (GBDT) to produce the final prediction in customer retention. While the CNN could learn feature patterns through non-linear relationship, it is hampered due to the adoption of convolution filters leading to uncertainty in data boundary partition and loss of useful edge points (Zhang et al., 2018).

Also, MLP model was used by Amatare & Ojo (2020) to achieve feature extraction and prediction, however, MLP has a shallow structure, which may result to minimal feature pattern learning. Designing DL models, therefore, requires careful consideration, aiming not only to extract features but also to extract features that can enhance performance robustly. This becomes particularly important if DL models are to maintain a competitive edge over traditional ML approaches, especially in scenarios involving non-image data. Recent concepts, including residual net (Allen-Zhu & Li, 2019; He et al., 2020) and attention mechanism (Wu, 2022) are promising techniques to enhance the performance of CNN, while a tree-structured concept (Balestriero, 2017; Yang et al., 2018) was proposed to improve MLP model. Developing a hybrid model, combining enhanced CNN and MLP based on these concepts are areas not widely investigated in CRP model in the telecom.

1.5 Research Questions

The following are the research questions that this research will address:

1. What approach can solve the imbalance data in the telecom CRP that is data-centric and computationally less expensive to build cost-enabled DL models?
2. What approach can improve the LR update as an essential parameter for improving the cost-enabled DL faster convergence and better generalisation?
3. What approach can sufficiently improve feature engineering to gain overall performance of the cost-enabled DL classifier in identifying customers leaving one service provider to another?

1.6 Research Objectives

Based on the issues highlighted in the problem statement, the aim of this research is to predict classifier in identifying customers leaving one service provider to another. The specific objectives are:

1. To design a cost-metric approach called Class-Imbalance-Ratio-Weight (CIRW) to solve the imbalance problem that easily integrates with DL models with less computational cost to form the cost-enabled hybrid DL.
2. To improve the existing CARL called Period-Shift Cosine Annealing LR (ps-CALR) method to optimise the cost-enabled hybrid DL models training process.
3. To design a hybrid DL classifier, combining RCA and TSM to learn improved features to gain overall predictive performance while integrating CIRW and ps-CALR.

1.7 Research Scope

The research study aims to develop cost-enabled hybrid deep networks integrating CIRW and ps-CALR to investigate customer retention in the telecom industry. The initial setup of the study entails sourcing the datasets and preprocessing. The proposed approach to solve the imbalance problem in the datasets uses the basic principle of class imbalance ratio specific to each dataset. This approach allows setting weight specific to both the majority and minority class in each dataset. The proposed CIRW compares only with conventional data resampling methods. In this regard, oversampling methods were only considered because the sample sizes of the datasets were not large enough for downsampling to be applied. Moreover, the DL model requires much data to perform well.

In another vein, carefully choosing hyperparameters, especially LR with optimizers, imparts deep network model performance. Hence, the research proposed ps-CALR to improve existing CALR. The proposed LR scheduling technique compares only with existing CARL and not necessarily with other LR schedule policies.

In addition, residual net, attention and tree-wise mechanisms were used to formulate the proposed hybrid deep network. The residual block and attention were used to enhance the convolutional block (1D-CNN), while a tree-structured concept serves as improvement from the MLP model. These concepts strengthen the feature extraction and performance of the overall DL model. The feature extraction of each part is concatenated to provide a final prediction.

The hybrid deep network incorporates both the proposed CIRW and ps-CALR techniques. The CIRW integration imparts cost sensitivity to handle the imbalance in the datasets, while the ps-CLALR integration facilitates quicker convergence and enhances generalization capabilities of the DL model.

The performance evaluation metrics considered in this research include F1-Score, AUC-Area under ROC (Receiver Operating Characteristics) and Geometric Mean (Gmean). No emphasis is placed on accuracy because of the imbalance datasets used. Lastly, Python programming language, TensorFlow, and Keras models are data pre-processing and model training tools used in the research.

1.8 Main Contributions

The following are significant contributions this thesis makes to the body of knowledge:

i. A CIRW method integrable in DL method to solve the imbalance data problem without additional computational cost and data symmetry loss. Thereby engendering on the ability to dynamically adjust the impact of different class instances based on the relative imbalance in the dataset, offering a fine-grained solution to the challenges of imbalanced datasets.

ii. A ps-CALR policy is designed to improve on the existing CALR to achieve faster convergence and better generalization of DL models through sufficient exploration-exploitation of loss region.

iii. An enhanced CNN block was created by incorporating a residual block and attention mechanism, resulting in the formation of an RCA model.

iv. An improved MLP block called TSM was modeled by introducing the decision tree concept to the MLP model.

v. A cost-enabled hybrid, combining RCA and TSM models is designed while integrating CIRW and ps-CALR for customer retention prediction. The hybridization demonstrates the ensemble of two deep neural networks to improve model performance.

1.9 Outlines of the Research

The structure of this research consists of six main chapters, which include:

Chapter 1: introduces the research background, motivation for the research, problem statements, research objectives, the scope of the research, and expected contributions of the research.

Chapter 2: presents the literature reviews on theoretical concepts of customer retention, discusses factors affecting churning, supervised learning, managing imbalance data, DL method, LR policies, and related research on CRP.

Chapter 3: presents the fundamental approaches and research methodology used in the research work related to the data science process. The chapter also discusses performance metrics for model evaluation and all the tools used in the research.

Chapter 4: highlights the specific techniques and frameworks proposed in the research, discussing the CIRW and ps-CALR approaches. Additionally, it delves into the components and concepts of the proposed cost-enabled hybrid DL architectural framework, focusing on the enhancement of CNN to form RCA and MLP to form TSM models

Chapter 5: presents the results and discussion of the proposed methods and various analyses conducted. It covers the evaluation of CIRW, comparing it with existing oversampling methods. The results of ps-CALR are also presented and assessed alongside existing CARL methods. Furthermore, the overall results of the proposed model are discussed and evaluated.

Chapter 6: presents the conclusion from the findings, contributions, and recommendations for the future works of this research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Everyone looks for space in the present digital economy, where technology is the fulcrum controlling everything. Businesses are constantly in a competitive manner, positioning campaign strategies to ensure existing customers are better satisfied. Discussing on these issues further, this chapter is organized as follows: Section 2.2 gives an overview of Customer Relationship Management (CRM) system by extension customer retention as an integral part of CRM is discussed in Section 2.3, while factors affecting churning discussed in Section 2.4. Section 2.5, 2.7, 2.8, and 2.9 focus on ML techniques with further discussion on imbalance learning, DL, LR optimization and model explainability. Related studies are the focus of Section 2.10, and the chapter summary is presented in Section 2.11.

2.2 Customer Relationship Management System

The constant dynamics in the business world, especially for service providers, is a source of concern in finding new ways of rebranding and repackaging their products and services in the competing market to meet customers' expectations and continuous relevance (Chagas et al., 2020). Hence, using technology to acquire, process, analyse, and communicate customer data (both current and prospective) in a fashion that develops interesting patterns to provide for timely and better service-oriented decision-making has become inevitable in the modern world. Therefore, Customer Relationship Management (CRM) focuses on a unique and cost-effective customer experience that uses data-driven strategies and concepts to improve business objectives. Four major categories of CRM identified by Al-Homery et al. (2019) include strategic, operational, analytical, and Collaborative.

While each category has supported the growth of CRM, attracting customers, learning about them, finding the most suitable way of serving them, and then using this knowledge to retain them while promoting profit maximisation are hallmarks of centred around analytical CRM (Singh et al., 2020). The key components of analytical CRM, as illustrated in Figure 2.1 include customer identification, customer attraction, customer retention, and last but not least, customer development (Al-Homery et al., 2019; Singh et al., 2020).

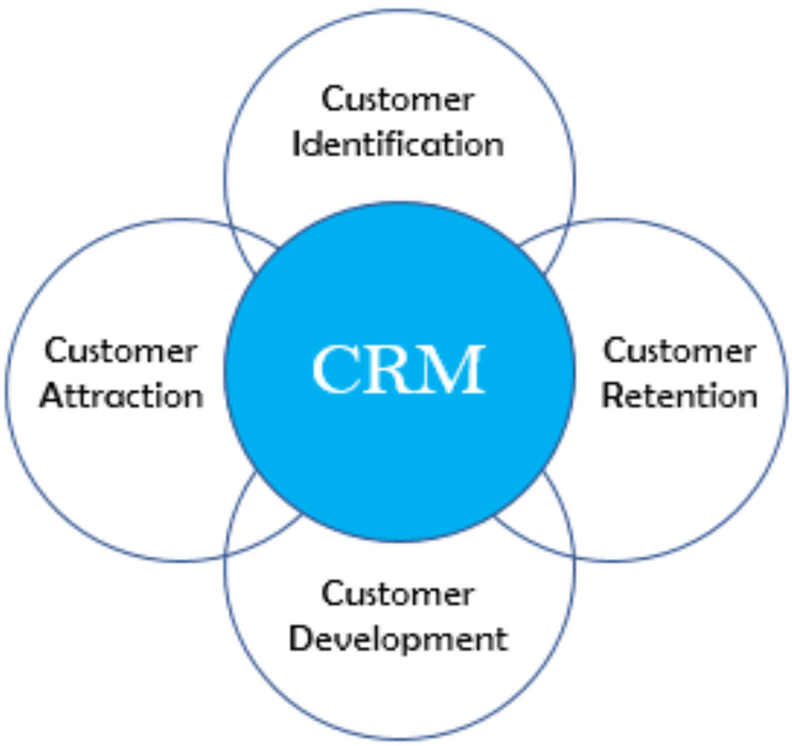

Figure 2.1 CRM and its four aspects in customer experience

Customer identification involves pinpointing potential customers and categorizing them into segments based on specific criteria (Alkhayrat et al., 2020). Once identified, customer attraction prioritizes these target customers by offering appealing resources tailored to their needs and interests(Lamrhari et al., 2022). Customer retention focuses on strategies and actions designed to meet and exceed customer expectations, ensuring their continued satisfaction and loyalty (Al-Mashraie et al., 2020). This can include loyalty programs, promotional offers, and efficient complaints management. Finally, customer development aims to enhance Customer Lifetime Value (CLV) by increasing transaction amounts, business value, and overall customer engagement in a systematic way (Karuppaiah & Palanisamy, 2021). Together, these components form a comprehensive approach to managing and nurturing customer relationships effectively.

Recent advances in CRM systems have significantly enhanced the impact on customer retention strategies. With the integration of artificial intelligence (AI) and machine learning (ML), CRM systems can now predict customer behaviour more accurately, allowing businesses to anticipate needs and tailor their offerings accordingly (Lamrhari et al., 2022; N. Singh et al., 2020). AI-powered chatbots and virtual assistants provide real-time support, improving customer satisfaction and reducing churn rates (Chatterjee et al., 2021; Lokuge et al., 2020). Furthermore, CRM systems now leverage big data analytics to process vast amounts of customer data, uncovering deeper insights into customer preferences and buying patterns (Lokuge et al., 2020). This allows for more personalized marketing campaigns and improved customer segmentation, ultimately leading to higher retention rates.

Additionally, advancements in CRM technology have facilitated omnichannel communication, enabling businesses to engage with customers seamlessly across multiple platforms, including social media, email, and mobile apps (Farmania et al., 2021; Lokuge et al., 2020). This cohesive approach ensures a consistent and personalized customer experience, regardless of the channel. Cloud-based CRM solutions have also made it easier for businesses to access and manage customer data from anywhere, enhancing flexibility and collaboration among teams (Chatterjee et al., 2021; Farmania et al., 2021; Lokuge et al., 2020). Moreover, the integration of Internet of Things (IoT) devices with CRM systems provides real-time data on product usage and customer interactions, allowing businesses to proactively address issues and improve service quality (Lokuge et al., 2020).

These technological advancements in CRM systems have transformed customer retention strategies by providing businesses with the tools to offer personalized, timely, and effective customer interactions. By leveraging AI, big data analytics, omnichannel communication, and IoT integration, businesses can enhance their understanding of customer needs, improve service delivery, and foster long-term customer loyalty. Consequently, modern CRM systems have become indispensable in the pursuit of sustainable competitive advantage and business growth.

2.3 Customer Retention

Customer retention encompasses the company's ability to meet and keep its present customers, often regarded as a critical component of successful CRM. CRP, which identifies customers with a high likelihood of attrition, is essential to customer retention. Firstly, a typical CRP model generalises the link between churn behaviour on the one hand. Secondly, based on historical data, customer attributes, and behaviour, it provides the company with data to make reasonable predictions about its customers' future behaviour and profitability (Amin et al., 2019). The antecedent benefits of effective CRP cannot be undermined. They include helping campaign policy-makers target marketing efforts cost-effectively and confining marketing efforts to a subset of consumers but with wider reachability of all customers with genuine reasons to leave. In addition, it reduces the need to acquire many new customers regularly, and maintenance of some sort of profit maximisation level is attainable.

The general and acceptable hypothesis is that acquiring new consumers incurs more expenses than maintaining the existing customer base. On the other hand, Amin et al. (2015) noted that customer churn behaviour unattended has an inevitable side effect on the company, including revenue loss (placing telecom at 21% among companies with the highest churn rate shown in Figure 2.2), risk of gaining new customers etc.

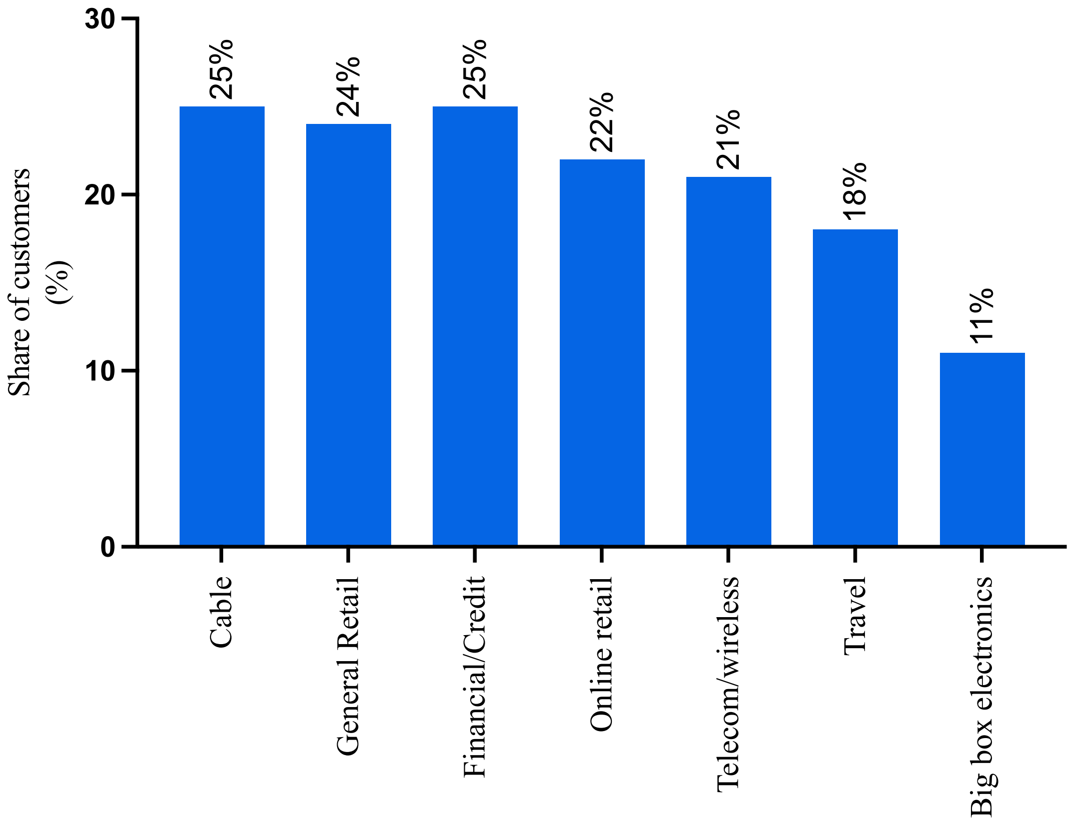

Figure 2.2 Churn rate in selected companies in the US (Statista, 2023)

With due consideration, telecom providers commonly classify customers into distinct groups, employing criteria such as billing status, mode of tariff payment, or churn behaviour (Swetha & Dayananda, 2020). The classification based on billing or payment mode encompasses professional customers (type 1), post-paid customers (type 2), and pre-paid customers (type 3). Conversely, classification based on churn behaviour includes active churners, passive churners, and rationale (silent) churners.

Understanding the intricacies of customer behaviour becomes imperative for an effective customer retention prediction using ML techniques. This knowledge is the foundation for developing robust models that anticipate and address different customer segments' diverse needs and preferences, ultimately contributing to enhanced customer satisfaction and loyalty (Amin et al., 2015).

2.4 Factors Affecting Customer Churning

Because CRP relies on the customer's behavioural pattern, certain attributes are believed to influence this pattern. Customer information is structured to include explanatory variables and the target class. The customer attributes provide insights into "why" customers churn. Hence, customer churning stems from customer satisfaction or dissatisfaction (Li et al., 2022).

There are many definitions and perspectives of customer satisfaction in the literature. According to marketing theorists, it is a shift in attitude, evaluation, and emotional reaction demonstrated by the customer in purchasing or repurchasing products or services following his expectations and perceived service rendered (Leninkumar, 2017).

Moreover, customer satisfaction lies in two other concepts: customer loyalty and customer trust. Loyalty is a function of attitude and behaviour (i.e., customer satisfaction). Customer loyalty is essential in gaining a competitive advantage over other service providers in a highly competitive and dynamic environment. Loyal consumers do not switch.