CHAPTER I

INTRODUCTION

1.1 Background

The Indonesian banking sector has been reshaped by digitization, mirroring a worldwide trend towards online financial transactions and services. This digital evolution is attributed to factors such as the COVID-19 pandemic, the rise in online shopping, and the advent of innovative digital payment systems [1], [2]. BCD BANK, a regional development bank primarily operating in North Sulawesi, is among the institutions navigating this transformation. Originally serving civil servants and providing credit services to local governments and their employees, BCD BANK has broadened its offerings to adapt to the changing banking environment and consumer preferences.

As digital banking transactions continue to flourish, BCD BANK, along with many regional banks, faces the challenge of efficiently handling and leveraging the vast quantities of data these transactions generate. In the digital age, understanding customer behavior – often referred to as "Know Your Customer" (KYC) – is critical for business success. This knowledge allows institutions to customize their offerings to customer needs, thus enhancing their marketing efforts and providing meaningful analytics [3] – [6]

In this process, customer segmentation plays a pivotal role. A widely used marketing framework for this purpose is the Segmentation, Targeting, and Positioning (STP) framework. It involves grouping potential or existing customers into segments based on similar characteristics, whether they be demographic,

geographic, psychographic, or behavior-based. For BCD BANK, this segmentation is primarily achieved through savings accounts, which offer features like Mobile Applications, Short Message/Messaging Service (SMS) Applications, and Automated Teller Machines (ATM).

The Recency, Frequency, Monetary (RFM) model, based on the Pareto principle (also known as the 80/20 rule), is a prevalent method for customer segmentation in banking [7] – [11]. This principle suggests that 80% of an organization's profits often come from 20% of its customers, reinforcing the need for effective customer segmentation [4]. The RFM model evaluates three variables: the recency of transactions, the frequency of transactions, and the transaction amount (monetary). Some researchers have added an extra variable, balance, to the RFM model, forming the RFM+B model [7], [12].

However, while traditional RFM analysis can offer substantial insights, it may not fully discern the complex relationships and patterns within its variables. Machine learning techniques, particularly the K-Means clustering algorithm, can supplement RFM analysis to provide more nuanced customer segmentation [13] – [15]. This algorithm is known for its efficiency and understandability, making it particularly effective when the number of clusters is known and working well with large datasets.

Bank Indonesia report in 2021, around 72% of the total banking transactions were conducted digitally, signifying the urgency for banks like BCD BANK to adopt advanced data management practices and enhance their service delivery. By integrating RFM analysis with K-Means clustering, banks can generate a more

granular and insightful segmentation of their customer base, which can subsequently guide targeted marketing strategies and improve customer relationship management.

Adopting sophisticated data analysis techniques like the integration of RFM analysis and K-Means clustering provides an opportunity for institutions like BCD BANK to optimize their data management practices. This becomes increasingly crucial given the growth of digital transactions and the need for banks to evolve with their customers' changing behaviors and preferences.

1.2 Problem Identification

Within the rapidly evolving digital banking landscape in Indonesia, regional banks like BCD BANK are challenged by the need for sophisticated customer segmentation to stay competitive. Currently, BCD BANK relies mainly on traditional segmentation practices centered on basic demographics, geographical locations, and the identification of the top 20 customers. These strategies, though satisfying regulatory reporting requirements, fail to deeply understand individual customers' transaction behaviors and loyalty drivers.

In this work, we aim to bridge this gap in customer segmentation that could potentially limit the competitive edge of BCD BANK amidst the rapid digitization of banking services. We propose the exploration and application of an advanced segmentation model, integrating RFM analysis and K-means clustering, to help improve customer relationship management and accurately identify and categorize loyal and frequent customers. This approach extends beyond basic demographic information and regulatory reporting, allowing us to develop more targeted and effective customer retention strategies.

1.3 Scope and Limitations

This study's primary scope involves the utilization of machine learning techniques, specifically the K-means clustering algorithm, combined with RFM analysis to refine customer segmentation. The chosen dataset comprises transactional data from the customer database of BCD Bank over the first half-year of 2022. The principal steps executed in this study are:

1. Data Acquisition and Preprocessing
2. RFM Model Adaptation
3. Exploratory Data Analysis (EDA)
4. Data Standardization (MinMax Scaler)
5. RFM Score Computation
6. Optimal Cluster Determination
7. Data Distribution Examination
8. K-Means Clustering Algorithm
9. Cluster Validation

While our methodology is comprehensive, it does not include certain aspects due to limitations in resources, time, and scope. Here are the limitations of this study:

1. **Data Quality and Availability**: The analysis strongly depends on the quality and availability of transaction data. Inaccurate or incomplete data could lead to less accurate segmentation results.
2. **Static Model**: Our RFM model, by itself, is static and doesn't account for evolving customer behaviors over time.

3. **Variable Interpretation**: The interpretation of the RFM variables can vary widely depending on the business context, and our research does not incorporate all possible interpretations.
4. **Lack of Personalization**: The RFM model provides a broad-brush approach to segmentation, but it may not account for individual differences and preferences among customers. This study is out of scope for personalization level that may require additional data and more complex analytical techniques.
5. **Implementation Complexity**: The implementation of machine learning techniques for customer segmentation may demand a significant upskilling of the workforce and investment in technology infrastructure, which our study does not cover.

The above limitations should be taken into account when interpreting the results of our research.

1.4 Problem Definition

Given the limitations stated above, this study revolves around the following consolidated research questions:

1. How can an advanced customer segmentation model, leveraging the RFM model and K-means clustering, provide a nuanced segmentation of BCD Bank's customer base?
2. How can this model delineate customer segments such as 'potential', 'passer-by', 'prime', and 'at-risk'?

3. How can the effectiveness of the developed customer segmentation model be measured and evaluated, particularly regarding its capability to identify and categorize loyal and frequent customers at BCD Bank?

These research questions aim to examine the possibility of transcending beyond basic demographic information and regulatory reporting requirements by developing a customer segmentation model that effectively utilizes RFM analysis and K-means clustering.

1.5 Research Purpose

The purpose of this research is threefold:

1. To adapt the Recency, Frequency, and Monetary (RFM) model to customer transaction data and utilize the K-Means clustering algorithm for segmentation of BCD Bank's customer base.
2. To identify the optimal number of customer segments that present within the Bank's consumer base.
3. To interpret the characteristics of the resulting customer segments and provide insights that could guide strategic decision-making for better customer engagement and retention in BCD Bank.

The realization of these purposes will be fully demonstrated in Chapter 4, where we apply our methodology to BCD Bank's data, determine the optimal number of clusters, and draw actionable insights from the resulting customer segments.

1.6 Outline of the Thesis

This research consists of 5 chapters. Every chapter elaborate its content each with specific purpose. Herewith the systematics of every chapter:

Chapter I Introduction. This chapter provides an introduction to the research topic. It starts with the background of the study, highlighting the importance of customer segmentation in digital banking services, specifically within the context of an Indonesian local bank. The chapter then delves into the problem statement, explaining the limitations of the current customer segmentation practices at BCD Bank. Subsequently, it formulates the research problem, posing critical questions to address the existing gap in customer segmentation methods that BCD Bank never done before. The research purposes are then presented, establishing the objectives this study aims to achieve. The chapter concludes with an outline of the subsequent chapters.

Chapter II Literature Review and Theoretical Framework. This chapter presents an in-depth review of existing literature on customer segmentation, the RFM model, and the K-means clustering algorithm. It identifies the gaps in the current literature and positions this research within the broader academic discourse. Further, it introduces and details the theoretical framework underlying the proposed advanced customer segmentation model, justifying the choice of RFM analysis and K-means clustering.

Chapter III Research Methodology. This chapter outlines the research methodology, detailing the data collection and data analysis processes. It introduces the dataset used for the research and elaborates on the application of the RFM model

and K-means clustering for customer segmentation. The chapter also explains the process of measuring and evaluating the effectiveness of the developed model.

Chapter IV Result and Discussion. This chapter presents the results derived from the analysis and discusses their implications. It evaluates the quality of customer segmentation achieved through the new model and assesses its capability in identifying and categorizing different types of customers. The chapter also contrasts the findings with existing literature, explaining any similarities or discrepancies.

Chapter V Conclusions, Implications, and Future Research. The final chapter provides a summary of the research findings, their implications, and the contribution of this study to the field of customer segmentation in digital banking services. It also discusses the potential practical implications of the developed model for BCD Bank and similar institutions. Finally, the chapter addresses the limitations of the study and suggests avenues for future research, fostering the possibility of further advancement in this field..