International Journal of Mathematical, Engineering and Management Sciences
Vol. 10, No. 3, 729-753, 2025
https://doi.org/10.33889/IJMEMS.2024.10.3.037

Ram Arti Publishers

Data Monetization Through Cross Industry Collaboration in Retail Banking

Sandeep Dey
School of Management Sciences,
Maulana Abul Kalam Azad University of Technology, Kolkata, West Bengal, India.
Corresponding author: sandeep.dey2@ltimindtree.com

Prasun Das
Statistical Quality Control & Operations Research,
Indian Statistical Institute, Kolkata, West Bengal, India.
E-mail: prasun@isical.ac.in

Indranil Mukherjee
School of Management Sciences,
Maulana Abul Kalam Azad University of Technology, Kolkata, West Bengal, India.
E-mail: indranil.mukherjee@makautwb.ac.in

(Received on July 18, 2024; Revised on September 30, 2024 & November 3, 2024; Accepted on November 5, 2024)

Abstract
This study investigates how data sharing between retail banks and e-commerce platforms, facilitated by data monetization, can improve customer experience in banking. Recognizing that most banking customers also utilize e-commerce services, the research explores how collaboration can benefit both parties. By analyzing customer data from both industries, the study develops propensity models to achieve market penetration and enhanced customer satisfaction. These models help identify high-potential customer segments for targeted product and service offerings. Conversely, e-commerce platforms can leverage banking data to target credit card promotions to customers with a history of high spending or large credit limits. This collaboration allows both industries to personalize their offerings and recommendations, ultimately leading to a more positive customer experience. The study proposes a novel framework for customer experience improvement through this collaboration. This framework utilizes three key pillars: portfolio segmentation, lead generation through e-commerce attribute propensity modeling, and banking attribute propensity modeling. By segmenting customers based on shared characteristics and predicting behavior based on specific data sets, the framework allows both industries to identify valuable leads and personalize their offerings, fostering customer acquisition and satisfaction. This research, focusing on a B2C collaboration approach, contributes valuable insights to a less-explored area within customer experience research.

Keywords- Cross industry, Customer satisfaction, Retail bank, Ecommerce, Customer experience.

1. Introduction

The biggest challenge before any service industry today is to provide enhanced satisfaction to its customers and thereby improve its Customer Lifetime Value (CLV). Retail banks are striving to achieve this by tapping their own customer database and using tools of Business Analytics (BA) and Business Intelligence (BI) to identify the issues which need to be addressed to reach this goal. However, these customers are not exclusive to the retail banks. Rather, there is a significant overlap of customers with other service industries such as telecom, food chain, e-commerce, retail FMCG as well as other banks. For example, if a bank wants to promote various online payments, then transaction data may not be sufficient because the customer may not use his / her account with this bank for any kind of online transactions. However, if the bank had data about the various online transactions that the customer is performing through other bank accounts then the bank can launch a customized campaign targeted towards the particular customer group. This is generically true for customers intercepted by other industries as well. To help improve the level of their satisfaction,

retail banks need to acquire data about the transactional and behavioral patterns of the customers who are tagged with other industries. The analysis of these data obtained through cross industry collaboration will help to highlight new dimensions about the customers which will never be possible if retail banks confine themselves to their own data bases only.Some of the topics considered in this connection include exploring the unified payment system and extracting all the connections that provide the attributes for potential collaboration such as banking with Food chain, transportation (OLA / UBER), communication (Vodafone Idea / Airtel / Jio), Fashion (Shoppers stop / Pantaloons), Groceries (Big Bazar / Spencer's / Reliance Fresh) etc.The advent of new business concepts using various digital techniques and e-commerce websites routinely generate huge amounts of customer data. Hence the cross-industry collaboration will enable the application of data monetization (Hanafizadeh and Nik, 2020; Kulhanek, 2019) techniques, viz. leveraging the data to generate value for the customers and the organization, adding new services to existing ones, developing new business and revenue models etc. thereby moving forward from data driven insights to the actionable. It is this idea which sets the context for the present work.

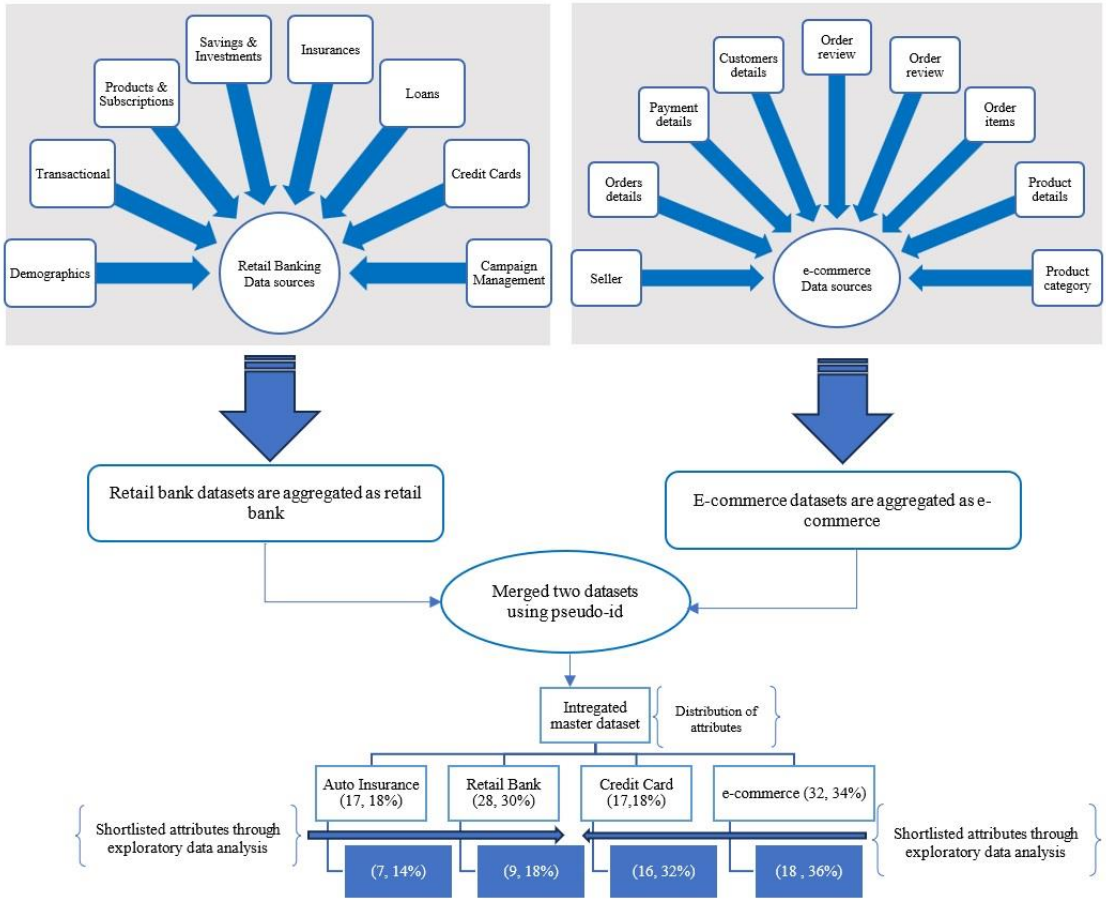

Figure 1. Data flow diagram.

The present work will focus on improvement of customer experience in retail banking and e-commerce domains by subjecting the data obtained through cross industry collaboration to data monetization tools and processes. The parameters for customer experience in e-commerce and retail banking would be different from each other. Hence, a convolution of the parameters and analysis of their inter-dependencies is required

which will address the issue of better customer experience, thereby leading to the development of an integrated framework.

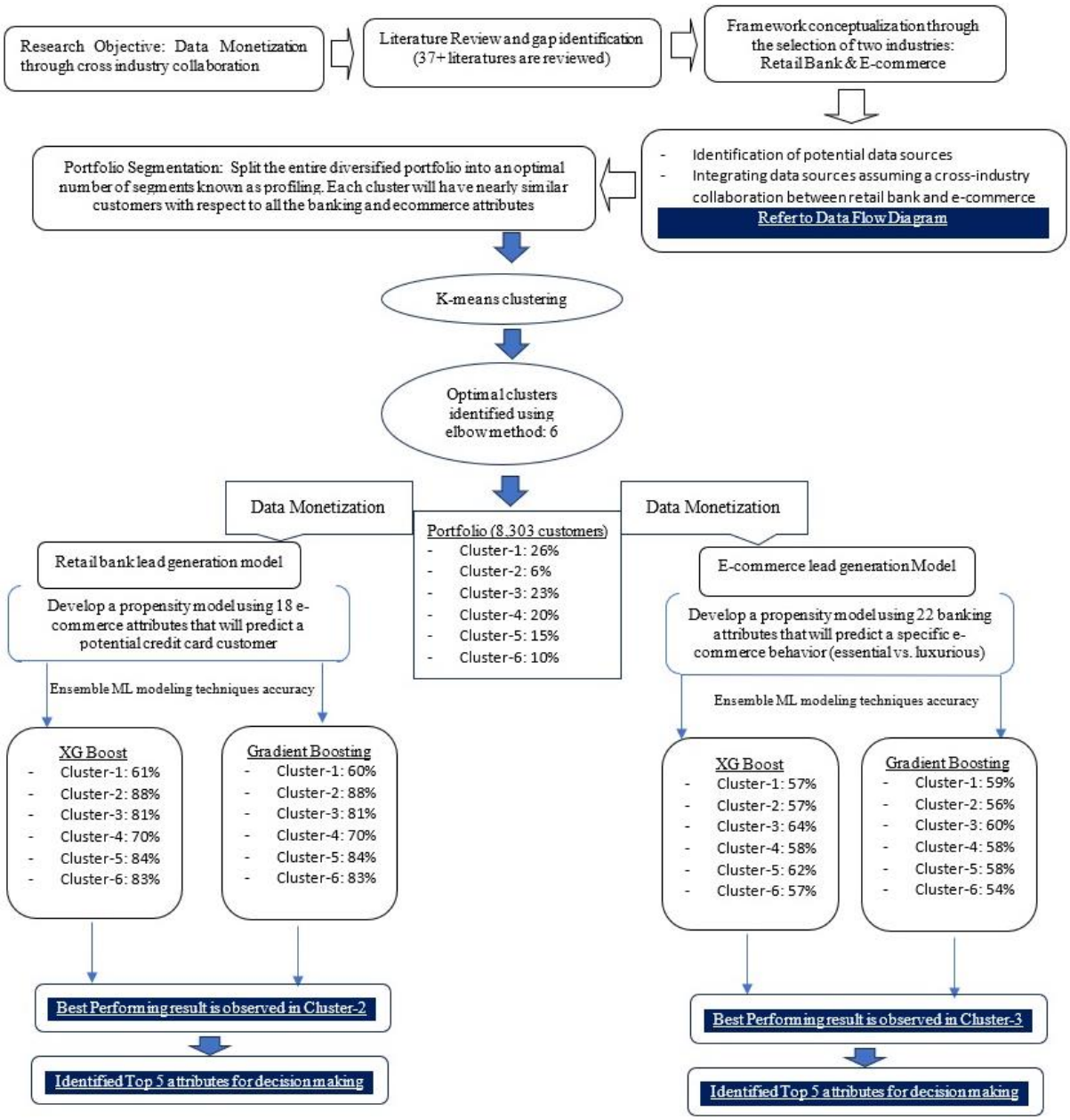

Figure 2. Process flow diagram.

The conceptualization of the framework is developed through three phases. During the first phase, a synthetic database is prepared with the assumption that cross-industry collaboration has occurred between a retail bank and an e-commerce company. The actual transaction data from both industries are extracted and merged based on pseudo-IDs. In the second phase, a portfolio segmentation using unsupervised algorithm is performed. The purpose of this segmentation is to identify and classify the entire portfolio into

multiple segments such that in each of the segments, all the customers are similar in nature with respect to all the attributes such as demographic, transactional, behavioral, social etc. During the third phase, two propensity models are developed. The first propensity model is developed for retail banks towards the banking product propensity whereas the second propensity model is for the e-commerce product-category propensity. The purpose of this model building exercise is to help both the retail banks and e-commerce companies in their marketing strategies and for developing marketing campaigns.

Currently some of the leading banks do have such collaboration such as co-branded cards with big retailers, but these may not be sufficient. However, this exercise will enable the retail banks to acquire 360^0 information about their customer base and enable them to work out strategies in the short and long term for better customer experience and management. The arrangement will benefit the retail banks as well as the collaborating industries. The framework will be distributed in multiple components such as inter and intra disciplinary communications. The implementation of Industry 4.0 is essential for effective process setup, planning, and preparation of production scenarios, as it requires collaboration between experts from various fields. This approach involves a theoretical framework for establishing inter-model relationships and visualizing them. The study is detailed using a Data Flow Diagram (**Figure 1**) and a Process Flow Diagram (**Figure 2**), illustrating the ideation and execution of the framework that demonstrates how data monetization can be achieved through cross-industry collaboration.

The rest of the paper is organized as follows. Section 2 discusses the literature review and the gaps in the existing literature. Section 3 presents a detailed account of the data preparation including identification of the source of data, synthesis and aggregation of the data, data cleaning and setting up of the workable database. Section 4 summarizes the exploratory data analysis. Section 5 focuses on model building and includes portfolio segmentation using K-means clustering as well as development of the propensity models. Section 6 discusses the results obtained while Section 7 deals with the managerial implications of the study. Section 8 draws broad conclusions and points out the directions in which the study may be extended in future.

2. Review of Literature

In this Section, we review the literature on various cross industry collaboration as well as customer satisfaction and e-commerce applications and processes. One such collaboration is between telecommunications and Internet of Things (IOT). Saragih et al. (2021) has studied cross industry collaboration between telecom and IOT through agile project management. In this study, a comprehensive strategic framework called EA (Enterprise Architecture) is developed, encompassing company analysis, design, and solution implementation. The study focuses on how to perform a successful cross industry collaboration from the operational point of view which is extremely important for all organizations. After successful implementation of cross industry collaboration, the next task is to measure customer satisfaction. Gilbert and Veloutsou (2006) worked on cross industry comparison of customer satisfaction. In their study the authors attempted to identify the empirically obtained core characteristics of customer satisfaction rating across six industries based on the rating of the respondents. This work provides an insight to the customer satisfaction perceived by the customers of other industries. Although this is not exactly the customer satisfaction achieved from cross-industry collaboration, yet it helps to understand the customers from the perspectives of both the industries involved in the collaboration. Lemon and Verhoef (2016) emphasized the importance of customer experience and customer journey over time for firms. Their study aimed to develop a robust understanding of these concepts in the current era of complex customer behavior. To achieve this, the authors analyzed existing definitions and conceptualizations of customer experience, offering a historical perspective on its roots within the marketing domain. This analysis points towards a path for measuring customer satisfaction.

In the context of customer experience measurement and monitoring, the products and services offered by the respective organizations play a crucial role. Raja et al. (2013) found that complex products such as manufacturing equipment always needed maintenance and repair. They also analyzed customers' views on integrated products and services and the value-in-use derived from such offering. They concluded that two key attributes viz. relational dynamics and access were found to have the maximum influence on customer satisfaction. This work provides a guidance towards identifying a potential opportunity in the space of products and services for cross industry collaboration. The next question is: what are the best practices for cross industry collaboration. In his research, Bader (2013) conducted a focused investigation into cross-industry innovation. Using a mixed methods approach, Bader explored Henkel's organizational learning process and the benefits it has yielded. This is particularly significant as Henkel, a global leader in both industrial and consumer business, has systematically employed cross-industry innovation to generate multiple radical innovations. Bader's work provides a new perspective on innovation that may be realized through cross industry collaboration.

The identification of Critical Success Factors (CSF) is the key for successful collaboration. Historically it is established that cross industry collaboration is very much successful for business to business (B2B) engagements. The critical success factors identified from B2B can be leveraged in B2C industries. Eid et al. (2002) identified few B2B critical success factors in the context of cross industry association. The authors observed that market entry and communication via the internet have strongly influenced the dynamics and traditional processes prevalent in B2B commerce. The authors discussed the difficulties resulting from this new trend. They presented a comprehensive review and identified 21 critical success factors applicable to most of the B2B associations. Ondrus and Pigneur (2007) have explored a completely different dimension of cross industry collaboration. The study conducted by Jupp (2016) focused on cross-industry preferences for mobile payments in Switzerland. The authors found that the industry favored card-based solutions over mobile phone-based solutions. Additionally, Jupp's research delved into cross-industry learning within the context of Product Lifecycle Management (PLM) and Building Information Modelling (BIM). This work offers a typology for comparing PLM and BIM, allowing for discussion and association of findings from existing studies on concepts, objectives, methods, and supporting technologies. The author concluded that significant learning can be achieved through cross-industry collaboration and is not confined to any specific industry vertical.

The study by Jamali (2007) examined customer satisfaction within the context of private-public partnerships, illustrating that cross-industry collaboration extends beyond different industries to also encompass collaborations between private and public organizations. The research aimed to present the outcomes of a unique satisfaction survey in the context of a new public-private partnership (PPP) within the Lebanese postal sector, emphasizing typically disregarded connections between PPPs, quality management, and customer satisfaction. The findings indicate a satisfactory level of satisfaction with the quality of services provided through the particular PPP under review but varied results regarding the influence of consumer characteristics on satisfaction ratings. Hence, it can be concluded that collaboration may not always lead to higher customer satisfaction. While there is a lot of focus on the collaboration process, identification of critical success factors and measurement of enhanced customer satisfaction, it is equally important to put stress on innovation. Bellingkrodt and Wallenburg (2015) conducted an analysis of the impact of customer relations on innovativeness and customer satisfaction within various service industries. The study aimed to offer insights into the similarities and differences in service innovation across sectors, specifically comparing logistics service providers and IT service providers (SPs). The research revealed that both types of service providers experience positive effects on innovativeness and customer satisfaction when maintaining close relationships with their customers. Francis (2024) dissertation investigates data monetization, turning data into profit, sharing data across industries, like retailers with

weather services, and unlocking new opportunities. This collaboration yields valuable insights that would not be possible alone, ultimately leading to improved products, services, and increased profit from data. In similar context, Sarkar and Sarkar (2019) examines data monetization for Indian telecommunication companies. Data is now their main source of revenue. The study emphasizes how telecommunication companies can leverage data monetization strategies and cross-industry collaboration in the digital age to maximize profits. This collaboration creates new business models within the digital ecosystem. In the context of Industry 4.0, Luo and Zahra (2023) highlight data monetization through cross-industry collaboration as an important aspect of Industry 4.0. Industry 4.0 is a major transformation impacting businesses globally. This collaboration allows multinational companies (MNEs) to leverage data for new opportunities in digitalization, intelligence, and innovation. Olaleye and Adusei (2024) explore data productization in the booming data economy. They identify the challenge of utilizing vast amounts of "dark data." Their research suggests that cross-industry collaboration is the key to unlocking the potential of this data for insights and improved decision-making. By working together, businesses can leverage data to enhance customer experiences and gain a competitive edge. Ofulue and Benyoucef (2024) investigated the potential of data monetization to enhance customer experience. Their literature review explored how cross-industry data collaboration can be leveraged across businesses. The study offers a three-pronged contribution: firstly, by expanding on existing frameworks, they provide a clearer picture of data monetization models; secondly, the authors categorize key topics and trends in data monetization; and finally, they identify practical applications for businesses and propose directions for future research on how data collaboration can improve customer experience. Ritala et al. (2024) explores how businesses can collaborate across industries (cross-industry collaboration) to use data for profit in B2B markets. They found companies often struggle to turn their data into sellable products or services. The study suggests how collaboration can unlock new value propositions for customers, like selling data directly, embedding data analysis in products, or offering data-driven services. This collaboration allows businesses to leverage each other's data to improve customer experience and gain a competitive advantage. The collaboration between retail banks and e-commerce entities is an increasingly prevalent trend, driven by the ongoing digital transformation of the financial services sector. This partnership seeks to provide innovative financial solutions, including digital payment options, seamlessly integrated banking services within e-commerce platforms, and personalized financial management tools. The objective of these alliances is to offer better customer experience, streamline online transactions, and capitalize on the growing popularity of online shopping. This collaborative effort includes the integration of banking services into well-known e-commerce websites or applications. There is also a focus on developing co-branded credit cards that offer enticing rewards for purchases made through specific e-commerce platforms.

Rao (1999) explored the impact of the Internet on distribution channels and discusses its influence on retailing, banking, and the music industry. It aims to highlight managerial implications and imperatives arising from this transformation. Lee et al. (2003) distinguishes between basic and collaborative B2B e-commerce adoption. They argue that firms are unlikely to benefit significantly from basic B2B e-commerce, while the use of collaborative B2B networks with channel partners can lead to dramatic improvement in firm performance. The study suggests that the source of performance improvement in B2B electronic commerce is not solely the electronic linkage, but also the collaboration enabled by the electronic network, as revealed through a survey conducted in the grocery industry. Padmavathy et al. (2012) aims to create a comprehensive scale for assessing customer relationship management effectiveness (CRME) in Indian retail banks. By identifying key dimensions, bank managers can enhance CRM practices, leading to improved customer satisfaction, loyalty, and cross-selling opportunities. The study explores the relationship between CRM efforts and outcomes such as satisfaction, loyalty, and cross-buying. Zwass (2003) has observed that over the past decade, e-commerce has become a significant force in organizational and social activities. To fully understand and leverage its impact, e-commerce can be categorized into five key

domains: commerce, collaboration, communication, connection, and computation. These domains offer innovative opportunities for organizing marketplaces, creating novel products, collaborating with business partners, transforming processes, and delivering information-system services. Therefore, e-commerce serves as a technologically based foundation for business transformation and a meta disciplinary research field. Hånell et al. (2020) investigates how market factors (such as institutions, competition, and resources) influence the international strategies of online retailers. This work introduces a novel theoretical perspective, contributing valuable insights to the international retail literature. Specifically, it explores the pros and cons of leveraging partnerships to address challenges in international online retailing. Kurnia et al. (2015) explore how market factors influence the international strategies of online retailers, emphasizing the use of collaborative efforts in overcoming different challenges related to international online retailing. Beck et al. (2005) conducted research on the adoption of e-commerce technologies by small and medium-sized enterprises (SMEs) in developing countries, finding that readiness at the organizational, industrial, and national levels, as well as environmental concerns, influence adoption. The study revealed that e-commerce applications have matured across various countries and industry sectors, benefiting both large firms and SMEs. In Denmark, Germany, and the US, SMEs reported significant improvements in operational efficiency and market expansion through e-commerce implementation. Efficient e-commerce usage was found to require comprehensive solutions such as online procurement and Internet-based supply chain management. Zott et al. (2000) examined value creation strategies for e-commerce companies across industries and identified two main strategies for value creation: efficiency in business models and creating customer loyalty based on a survey of 30 European e-commerce companies. Tarafdar and Vaidya (2006) framework proposed analyzing factors influencing organizational inclination to use e-commerce technologies based on qualitative data from four financial firms in India. Chan et al. (2012) described how organizations use collaborative supply chain management (SCM) to gain a competitive edge with a focus on Small and Medium Enterprises (SMEs). Zhu (2014) study examined the impact of e-commerce capability and IT infrastructure on firm performance in the retail industry highlighting the importance of integrating front-end ecommerce capability with back-end IT infrastructure. The studies mentioned above highlight that while there are challenges accompanying these collaborations such as regulatory complexities relating to data sharing and privacy protection; there exist opportunities for banks to expand their reach beyond traditional banking through strategic digitization initiatives involving partnerships with successful e-commerce players.

Existing research often delves into technical aspects such as data sharing, model building, and system integration in cross-industry collaboration. However, there remains a critical gap in understanding the specific value proposition for each collaborating business. This study aims to move beyond technical details and explore the “why” of collaboration. By analyzing the underlying motivations, businesses can identify strategic advantages beyond mere operational efficiency. For instance, collaboration might enhance customer experiences, open new markets, or foster innovation. Understanding these value drivers is essential for successful cross-industry partnerships.

The research focuses on collaboration between a retail bank and an e-commerce platform. It examines how targeted recommendations, facilitated by data sharing and analysis, impact both entities. Quantifiable benefits include revenue growth, improved customer acquisition rates, and enhanced customer satisfaction. Additionally, the study emphasizes that data monetization is crucial but often an overlooked aspect. Data monetization involves strategically leveraging insights from shared data to create new revenue streams or optimize existing processes. By exploring these strategies, the present work aims to ensure a win-win situation for both the bank and the e-commerce platform while considering ethical data privacy concerns.

3. Data Preparation

The literature review is foundational in developing an understanding of existing knowledge and theories related to the improvement of customer experience through cross-industry collaboration. It provides valuable insights and findings that help identify gaps and shape the framework of the study. The next crucial step is data preparation, which involves collecting, organizing, and cleaning data to ensure its quality and alignment with existing knowledge. The insights gained from the literature review inform decisions during data collection and organization, ensuring that the analysis being undertaken is grounded in existing scholarship while also contributing new insights to the field. This interconnectedness allows for a comprehensive approach that integrates previous research into the handling and preparation of data for this specific study. The preparation of data comprises of source identification, data synthesis, data aggregation, data cleaning and preparing the workable database. Each of these steps will be executed as follows:

3.1 Source Identification

Retail banking customer data is extremely confidential information and is stored within Customer Relationship Management (CRM). Banks do not allow any third party to have access to this information. So, the data collection exercise has been done through research on various banking websites, world bank data, blogs, data science and machine learning forums such as Kaggle, GitHub, Medium etc. The idea behind this data collection exercise is to develop a synthetic data which will be in alignment with the business objectives of the organization. The retail banking customer value management data can broadly be classified into seven categories which are stored in separate data bases, viz. demographics, transactional, products & subscriptions, savings &investments, insurances, loans, credit cards, & campaign management. E-commerce have been considered as collaborative industries in this study. There are couple of data sets distributed across all the dimensions as mentioned above, which are essentially actual data from CRM systems but the data will be masked in order to maintain customer confidentiality.

The study incorporates data from 15 distinct sources, with 13 exclusively pertaining to retail banking, and one each from e-commerce and auto insurance. The details of these 15 data sources are shown in Annexure 1. These 13 sources of retail banking are categorized under various CRM segments, including demographics, campaigns, and information related to credit cards and personal loans. The data, sourced from Kaggle, has anonymized personal identifiers, ensuring accuracy and authenticity. The primary challenge lies in the disparate origins of the data, necessitating careful consolidation and integration. The genuineness of the transactions, demographic details, card, and campaign data mandates a strategic mapping to unify the dataset as if it originated from a single source.

3.2 Basis for Data Synthesis

All banking data will be integrated using common attributes from separate data tables into one large data table and, in a similar fashion, all e-commerce data tables will also be integrated respectively. All the data tables pertaining to respective industries can be combined or merged using either single primary key or composite primary key. The concept of pseudo_id is introduced where the primary key is not available. The primary key is the common attribute which is present in all the data sources that helps to integrate all the data sources. Pseudo_id is a random number which is assigned to every record in each dataset. This is the same set of numbers that are randomly distributed to all the data sets. Suppose data set A has 100 records then all the records are assigned one number between 1 to 100. In a similar way, if another dataset has B has 500 records, then each record will be assigned a random number between 1 to 500. Both datasets will have one common attribute pseudo_id. On the basis of this variable the two datasets will be merged and the integrated dataset will be validated using a data quality framework. The data quality framework will compute the correlation between attributes of the two datasets. It is possible that there are few attributes which are common between the two cross industry data sets. Hence, the difference in the values will be

computed among the common numeric attributes and for common categorical attributes, binary variables will be computed between banking and ecommerce attributes across all the records. The next step would be to compute the weighted average from all these common attributes. This needs to be computed for each record in the integrated data set. Finally, the total difference will be computed across all the rows and then divided by the number of records. This one single number will be achieved for each iteration. A simulation exercise has to be performed with a sizable number of iterations and then an optimal number will be selected as the final data set.

In the absence of common attributes, some features have been identified from the banking dataset that will have some correlation with the features arising from the e-commerce data set. The data set has to be consolidated and aggregated for each customer based on the identified features using pseudo id. The identified features will be standardized and normalized. The nonlinear correlation among the features between the two data sets will be computed. The calculation will be done for each customer in the randomly integrated dataset based on the pseudo-id. Finally, the total correlation will be computed across all the rows and then divided by the number of records. This single number will be achieved for each iteration. A simulation exercise is to be performed with a sizable number of iterations and then an optimal number will be selected as the final data set.

The study also employs Principal Component Analysis (PCA) to reduce the dimensions which have to be studied for computing the data quality. This is done in order to reduce the computation time and cost to measure the correlation between attributes in the combined dataset. This, however, comes as a trade-off between reduction in dimension and reduction in the amount of variability in the data that can be explained by the attributes. This might lead to higher values of nonlinear correlation metrics–MIC and distance correlation, which might not necessarily mean a simultaneous improvement in the combination of the cross-industry data.

3.3 Assumptions

It has been assumed that all data components are from the same organization. Hence, the concept of pseudo-id creation is implemented. With the help of the pseudo-id, communality has been established. This pseudo-id will be behaving as a primary key and will help to merge all of the 15 data sources.

3.4 Aggregation

All the individual datasets are aggregated into three major datasets as retail bank, e-commerce and auto insurance datasets. After performing data aggregation, e-commerce and auto insurance datasets are merged with the retail bank dataset. Pseudo-id of each customer is created in retail bank-commerce and auto insurance datasets and it helps to project the customer's banking behaviour onto the e-commerce business and vice-versa. The data aggregation is achieved through a 4-step process which is described below.

Step 1: Integration of ecommerce data sources

- Brazilian ecommerce dataset comprises of seller details, orders details, payment details, customers details, order review, geolocation, order items, product details, product category details.
- Order review data set is a large dataset and hence we extracted the important attributes and excluded review creation date and review answer timestamp from the data frame. There is a total of seven variables and two variables were excluded i.e., review_creation_date and review_answer_timestamp which implies the extracted attributes are review_id, order_id, review_score, review_comment_title, and review_comment_message from order review data.
- Order item data set consists of a total of six variables, out of which two variables were excluded and only four variables were taken into consideration, viz. order id, product id, price, and freight value.

- Product details dataset consists of total of nine variables; we excluded four variables i.e., product weight (gm), product length (cm), product height (cm), and product width (cm). The remaining five variables taken into consideration are product id, product category name, product name length, product description length, product photos quantity.
- Orders dataset consists of order id, customer id, order status, order purchase timestamp, order approved at, order delivered carrier date, order delivered customer date, order estimated delivery date. Out of these only order id, customer id and order status seem important and are hence considered.
- Payment dataset consists of variables such as order id, payment Sequential, payment type, payment instalments and payment value.
- **Stage 1:** At first merge order item dataset with the product details dataset using product id.
- **Stage 2:** Output from stage-1 is merged with the order review data using the key column order id.
- **Stage 3:** Output of stage-2 is merged with orders dataset using the key column order id.
- **Stage 4:** Output of stage-3 is merged with Payment_Dataset using order id. The integrated output from stage-4 is the final ecommerce dataset.

Step 2: Demographic data preparation

- For demographic attributes, we used IBM-Watson datasets in order to retrieve important features. In IBM-Watson dataset, we have a total of 24 variables namely, Customer, State, Customer Lifetime Value, Response, Coverage, Education, Effective to Date, Employment Status, Gender, Income, Location Code, Marital Status, Monthly Premium Auto, Months Since Last Claim, Months Since Policy Inception, Number of Open Complaint, Number of Policies, Policy Type, Policy, Renew Offer Type, Sales Channel, Total Claim Amount, Vehicle Class, and Vehicle Size.
- Out of these 24 variables in IBM-Watson dataset, only 12 variables are extracted i.e. Customer, Coverage, Total Claim Amount, Monthly Premium Auto, Customer Lifetime Value, Education, Employment Status, Gender, Income, Location Code, Marital Status, and Vehicle Class.

Step 3: Banking dataset preparation

- **Step 3a:** Utilizing Portuguese bank dataset:
  - Portuguese Bank dataset consists of 17 variables such as age, job, marital status, education, default, balance, housing, loan, day, month, duration, campaign, p days, previous, p outcome, response, responsibility.
  - Out of 17 variables mentioned above, only 8 variables are considered as age, job, marital status, education, default, balance, housing, and loan.

- **Step 3b:** Utilizing Credit Card dataset:
  - Credit Card dataset consists of total 18 variables namely, Customer id, Balance, Balance frequency, purchase, one-off purchases, instalments Purchase, Cash Advance, Purchases Frequency, One-off Purchase frequency, Purchase's instalment frequency, Cash advance frequency, Cash advance transaction, Purchases Transaction, Credit Limit, Payments, Minimum Payments, PRC Full Payments, and Tenure.

Step 4: Consolidated dataset preparation

- Merging ecommerce dataset, demographic dataset, Portuguese bank dataset and Credit Card datasets.
- In order to merge different datasets, one common attribute is required for merging operations. Hence the pseudo_id for each customer is introduced and we make a key column in each dataset which is named as Pseudo_id column.

- IBM-Watson dataset consists of total 9,134 unique observations (rows) with 12 selected variables. One of the columns is customer which represents unique id of each customer. This attribute is replaced with pseudo_id which has 9,134 unique pseudo_id of each customer.
- Portuguese Bank dataset consists of 45,211 unique observations and 8 selected variables. One more column is added as pseudo_id to Portuguese Bank dataset which contains 45,211 unique pseudo_id for each customer.
- Credit Card dataset consists of 9,134 unique observations and 18 attributes. One of the attributes is customer id which is replaced with the pseudo_id column that contains unique pseudo_id for each of the 9,134 customers.
- E-commerce dataset consists of 1,18,318 unique observations and 19 attributes. Here also, pseudo_id is added as an additional column which contains 1,18,318 unique pseudo_id for each observation.
- Among all the data sets, the minimum customer base is 9,134 hence, the final dataset after performing all the inner joins results in 9,134 customers with all the 93 attributes.

3.5 Cleaning for Quality

Post aggregation, the duplicate records are identified and removed. There are few records for which most of the attributes are null. All such records are dropped from the data set. All the numeric attributes are analysed and checked for outliers for subsequent removal. There are few numeric attributes, similar in nature, which have been standardized using normality check after performing the necessary log transformation. Also, the pseudo-id is renamed as customer-id. All the variables are converted into lower case and the space replaced with the underscore.

3.6 Workable Database

All the 16 data sources are aggregated in four stages. First, all retail bank datasets are aggregated, to make one dataset as retail bank (dataset–I). Then, all e-commerce and auto insurance datasets are aggregated separately to make one dataset as e-commerce (dataset-II) and auto insurance (dataset-III). Finally, on the basis of pseudo-id of each customer, all these three datasets are merged into one master dataset for model development.

The objective of this aggregation is to bring all the attributes of retail banking customers, e-commerce customers and auto insurance customers under one umbrella. The merged dataset contains all the essence of banking and ecommerce industries attributes. The final data comprises of information of 9134 customers.

The goal of data integration is to merge customer profiles that span multiple industries like retail banking, e-commerce, auto insurance, and lending to create a unified view. This involves combining transactional and behavioral data to form a complete, 360-degree customer profile. With this all-encompassing view, companies can design personalized value propositions. During portfolio segmentation and predictive modeling for identifying potential leads, customers unique to one sector are not included, rather we are focusing on those who are common across various industries.

4. Summary of Exploratory Data Analysis

The aggregated and consolidated dataset obtained in Section 3 has 9,134 customers and 93 attributes. An extensive exploratory data analysis was conducted on all 93 attributes, examining various aspects of data quality including missing values, extreme values, inconsistencies, and correlations between attributes. It is observed that Retail bank (account, demography, credit card, loan and insurance) constitutes 66% variables. E-commerce (cross industry) contributes to 34% variables. There is a total of 1,26,913 customers (retail bank and e commerce). There are only 9,134 customers which are common between retail bank and e-

commerce. There are a total of 93 variables distributed across various segments of retail bank, credit card, auto insurance and ecommerce (**Table 1**). Out of these 93 variables, 50 variables are shortlisted through due diligence and domain understanding for model development (**Table 1**). It can be noted that no variable is selected from campaign info category due to incompleteness and poor data quality.

Table 1. Overview of retail banking & e-commerce variables.

Business segment	Data segment	Number of variables	Variable (%)	Number of variables	Variable (%)
		All variables (93)		Selected variables (50)	
Insurance	Auto insurance	17	18%	7	14%
Credit card	Credit card info	17	18%	16	32%
E-commerce	E-commerce	32	34%	18	16%
Retail bank	Account info	4	4%	1	2%
Retail bank	Campaign info	16	17%	0	--
Retail bank	Credit history	1	1%	1	2%
Retail bank	Demography	4	4%	4	8%
Retail bank	Loan info	3	2%	3	6%
Grand total		**93**	**100%**	**50**	**100%**

The list of all the 93 variables and its selected subset consisting of 50 variables used in the study are shown in **Annexure-2**.

After completing the various stages of data management and data preparation, the diversified portfolios of 9,134 common customers have been obtained. All these customers will be used to develop propensity models both for retail banking and collaborative industries. The models thus developed will be the winning strategy for Cross industry collaboration.

5. Model Building

The exploratory data analysis section is instrumental in identifying key variables for the development of a comprehensive framework, which encompasses statistical analysis, segmentation, and predictive modeling. This section focuses on developing three models: the first model utilizes unsupervised algorithms to segment portfolios, while the second and third models predict banking and e-commerce behavior, respectively. The purpose, details of the models, along with their respective processes and diagnostics are elaborated below.

Brief Overview of the Models

(a) Portfolio Segmentation using unsupervised algorithms: *K*-means cluster analysis will be performed that will split the entire diversified portfolio into an optimal number of segments known as profiling. Each cluster will have nearly similar customers with respect to all the banking and ecommerce attributes.

(b) Develop a propensity model using e-commerce attributes only that will predict a specific banking behaviour: A predictive model will be developed using only e-commerce variables which will predict the purchase frequency using credit card. The list of predominant attributes will be identified which drives high frequency credit card purchase. This information will be cascaded to bank so that bank can roll out personalized campaigns for these customers. All these campaigns will attract the customers towards purchase using credit card. Bank will be able to generate leads and increase profitable customers from all of these campaigns.

(c) Develop a propensity model using banking attributes only that will predict a specific e-commerce behaviour: In this case, a predictive classification model will be developed using only banking attributes to predict e-commerce product category, which a typical banking customer might prefer. The list of predominant attributes will be identified which drives customer preference while buying e-commerce

products. This information will be cascaded to e-commerce company so that the e-commerce company can design customized campaigns for all these customers. All these campaigns will surely attract some of the banks' customers towards e-commerce product purchase. In this way, e-commerce company will be able to generate profitable leads which will have high propensity for conversion.

Details of the Models

Portfolio segmentation using *K*-means clustering:

The model is designed to perform cluster analysis on the entire customer portfolio, grouping customers into distinct, homogeneous clusters. The primary objective of this model is to classify new customers into the appropriate clusters based on their characteristics and utilize a predictive model tailored to each cluster to identify potential leads for targeted marketing efforts. This two-step approach ensures that new customers are accurately segmented, allowing for more effective and personalized lead generation strategies.

The objective of performing K-means clustering is to split the entire diversified portfolio into homogeneous groups so that the customers belonging to each group are consistent. The selection of K-means clustering is justified by its simplicity, efficiency, and ease of interpretation. It partitions the data into K distinct, non-overlapping clusters based on distance metrics, typically Euclidean distance. It can handle large datasets effectively, making it suitable for a variety of applications. K-means is faster than many other clustering algorithms, especially on large datasets, because it typically converges in a few iterations. The only drawback of the K-means algorithm is the number of clusters should be known (supervised) in advance. The other clustering approaches, like hierarchical clustering or density-based spatial clustering of applications with noise (DBSCAN), may be preferred when the data contains non-spherical clusters, clusters of varying sizes, or when the number of clusters is not known beforehand. These methods can also handle outliers better than K-means. But in the K-means framework development, outlier detection is already performed, number of optimal clusters is identified and data contains spherical clusters. Hence, the K-means cluster is applied.

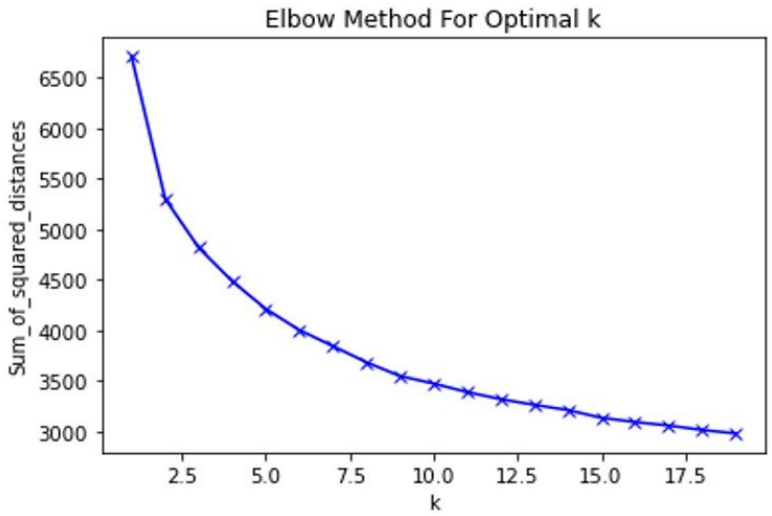

Figure 3. Optimal cluster identification.

K-means clustering algorithm incorporates 50 selected variables from various sectors, including retail banking, credit cards, e-commerce, and auto insurance, to conduct portfolio segmentation. These variables are integral to the cluster analysis process. To determine the most effective number of clusters, an elbow

curve is generated. By examining the inflection point on the elbow curve, the ideal number of clusters is established at six, as depicted in **Figure 3**. Consequently, the portfolio is divided into six uniform clusters. During the cluster analysis, 831 records were omitted, resulting in a distribution of the remaining 8,303 customer records across the six segments. The allocation of clusters is summarized in a table (**Table 2**) which outlines the segmentation across all six clusters.

Table 2. Cluster description.

Cluster name	Number of customers	Customers (%)
Cluster-1	2,147	26%
Cluster-2	498	6%
Cluster-3	1,874	23%
Cluster-4	1,666	20%
Cluster-5	1,251	15%
Cluster-6	867	10%
Grand total	8,303	100%

The explanation of cluster attributes and their significance are identified from cluster features (**Table 3**). Average purchase frequency is significantly high for cluster-5 followed by cluster-3 and cluster-6. Customers belonging to cluster-5 happen to make purchases mostly with full payments (one off purchase frequency). Average payments made by customers from cluster-5 happen to be dominating. Average income for cluster-4 is significantly high; however, the expense made by the customers is not so high. Credit limit & payments are directly proportional and this is also reflected from the cluster description.

Portfolio Insights are also identified by analysing the cluster features (**Table 3**). The average income for cluster-1 is significantly low as compared to other segments and the average payment is the second highest among all the six portfolio segments. The average payment and credit limit are minimum for cluster-2 among all the six portfolio segments; however, the average income of all the customers in this portfolio segment is not minimum. In cluster-5, purchase frequency is the highest among all the six portfolio segments. Average credit limit is highest; hence, it can also be concluded that the customers belonging to cluster-5 have high spending attitude.

Table 3. Cluster features.

Top 5 retail banking attributes define the portfolio segments					
Cluster name	**Purchase frequency**	**One-off purchase frequency**	**Payment**	**Income**	**Credit limit**
Cluster-1	0.13	0.07	1,619	14,763	4,220
Cluster-2	0.36	0.12	829	33,646	2,951
Cluster-3	0.87	0.13	1,546	36,254	4,133
Cluster-4	0.15	0.08	1,478	69,644	4,221
Cluster-5	0.92	0.82	3,421	39,510	6,799
Cluster-6	0.72	0.11	1,413	36,990	4,208

***Retail bank credit card lead generation model*:** The development of a propensity model is made focusing solely on e-commerce attributes to predict specific banking behaviors. This model aims to identify potential leads for retail banking services by analyzing online consumer patterns from e-commerce websites.

The present study aims to create a predictive model that utilizes cross-industry data from retail banking and e-commerce. The model's objective is to pinpoint potential customers who are likely to use a credit card for e-commerce transactions, based on their online transactional behaviour. The key dependent variable is binary, indicating whether a customer uses a credit card (1) or not (0) during an e-commerce transaction. A

total of 18 e-commerce-specific variables serve as independent variables in the model. The customer base for this model is shared between the retail bank and e-commerce platforms, which is why the use of a credit card during e-commerce transactions is the chosen dependent variable. All independent variables are exclusively related to e-commerce activities.

The data is divided into development and validation samples for each of the six segments. This approach helps prevent overfitting and helps to train the models on one subset and evaluate their performance on another. It is ensured that the data split is representative of the underlying distribution in each segment. In each segment, a separate model is built since different segments may exhibit unique patterns or relationships. However, consistent feature engineering techniques are used across all segments to maintain consistency. The model accuracy is monitored in the validation sample for all six segments. Accuracy is a common metric, but we also consider other evaluation metrics like precision, recall, F1-score, or area under the receiver operating characteristic curve (AUC-ROC). A strict monitoring and evaluation exercise is performed to ensure that the model does not overfit or underfit, especially since the segments have varying data characteristics. Among multiple machine learning algorithms, two top performers are identified: XGBoost and Gradient Boosting. Both XGBoost and Gradient Boosting are powerful ensemble methods that handle complex relationships and perform well in various scenarios.

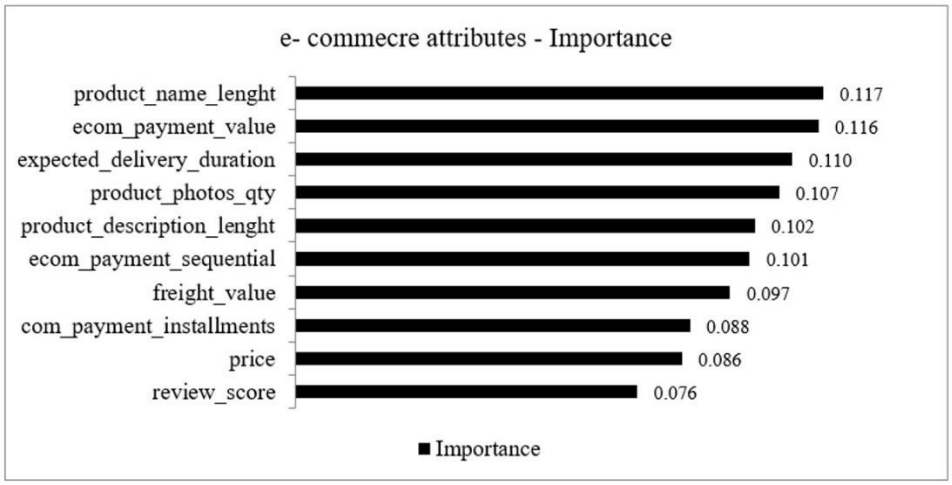

Figure 4. Importance of ecommerce attributes.

Table 4. Final model and portfolio segments.

Portfolio segment	XG Boost (Error %)		Gradient boosting (Error %)	
	Dev sample	Val sample	Dev sample	Val sample
Portfolio segment-1	27%	39%	25%	40%
Portfolio segment-2	11%	12%	10%	12%
Portfolio segment-3	16%	19%	16%	19%
Portfolio segment-4	23%	30%	22%	30%
Portfolio segment-5	15%	16%	14%	16%
Portfolio segment-6	14%	17%	14%	17%

In this study, the independent variables consist exclusively of e-commerce attributes. Among the six portfolio segments, accuracy played a crucial role in identifying the most suitable representative segment

and its corresponding predictive model for lead generation of potential credit card customers in a retail bank during e-commerce transactions. Notably, portfolio segments 2, 3, 5, and 6 consistently demonstrated good model accuracy (**Table 4**). However, the best-performing result was observed in portfolio segment-2 (**Table 4**). Consequently, deploying an individual model tailored to this specific segment is recommended. Additionally, the predictive model revealed the top 10 e-commerce attributes with significant importance (**Table 5** and **Figure 4**). Leveraging these features will be instrumental in developing a credit card customer acquisition strategy for any retail bank.

Table 5. Feature importance.

E-commerce attributes	Importance
Review_score	0.076
Price	0.086
Com_payment_installments	0.088
Freight_value	0.097
Ecom_payment_sequential	0.101
Product_description_length	0.102
Product_photos_qty	0.107
Expected_delivery_duration	0.110
Ecom_payment_value	0.116
Product_name_length	0.117

***E-commerce lead generation model*:** An e-commerce lead generation model is set up to predict customer preferences for specific products by analysing a retail bank's transactional and behavioural data. This work essentially uses banking information to identify potential leads for e-commerce businesses.

Table 6. Product category.

Serial number	Essential products	Luxurious / others
1.	Comfort home, home appliance, air conditioning, portable home oven & café, housewares, decoration / furniture, bed, table, bath	Agro industry and commerce, industry trade and business, office furniture
2.	Room furniture, furniture kitchen service area dinning and garden, garden tools	Construction tools, construction tools lighting, house construction
3.	Babies accessories, hygiene diapers	Automotive
4.	Signage & security, landline, telephony	Cool stuff, audio, arts
5.	Technical books, stationary store, accessory, bags	Sport leisure, game consoles, pc gamer, toys
6.	Electronics, informatics accessories	Beauty / health, perfumery
7.	Pet shop	Fashion handbags and accessories, gift watches
8.	Foods	

Table 7. Final model and portfolio segments.

Portfolio segment	Xg boost (Error %)		Gradient boosting (Error %)	
	Dev sample	Val sample	Dev sample	Val sample
Portfolio segment-1	81.49%	57.56%	97.34%	58.62%
Portfolio segment-2	92.33%	57.21%	98.17%	55.72%
Portfolio segment-3	84.09%	64.04%	98.00%	60.25%
Portfolio segment-4	79.78%	58.28%	97.40%	58.28%
Portfolio segment-5	89.60%	61.58%	97.85%	58.13%
Portfolio segment-6	77.44%	56.50%	97.83%	54.07%

A machine learning model is built to predict e-commerce product preferences for banking customers. It uses 32 attributes (9 banking, 7 auto insurance, and 16 credit card) to categorize potential purchases as

essential or luxurious and others (**Table 6**). This initial recommendation aims to guide customers in the right direction on the e-commerce website, although the challenge lies in capturing the variety of products customers might buy.

Table 8. Feature importance.

Retail banking attributes	Importance
Balance Frequency	0.018
Purchase transaction	0.019
One-off purchase frequency	0.022
Purchase frequency	0.022
Cash advance	0.024
Vehicle class	0.024
Education	0.025
Balance	0.026
Customer Lifetime value	0.027
Marital status	0.027
Total claim amount	0.030
Purcahses	0.031
Housing loan	0.032
Job category	0.032
Gender	0.033
Installment purcahses	0.033
Employment status	0.034
Monthly premium auto coverage	0.035
Cash advance frequency	0.037
PRC_ful_payment	0.044

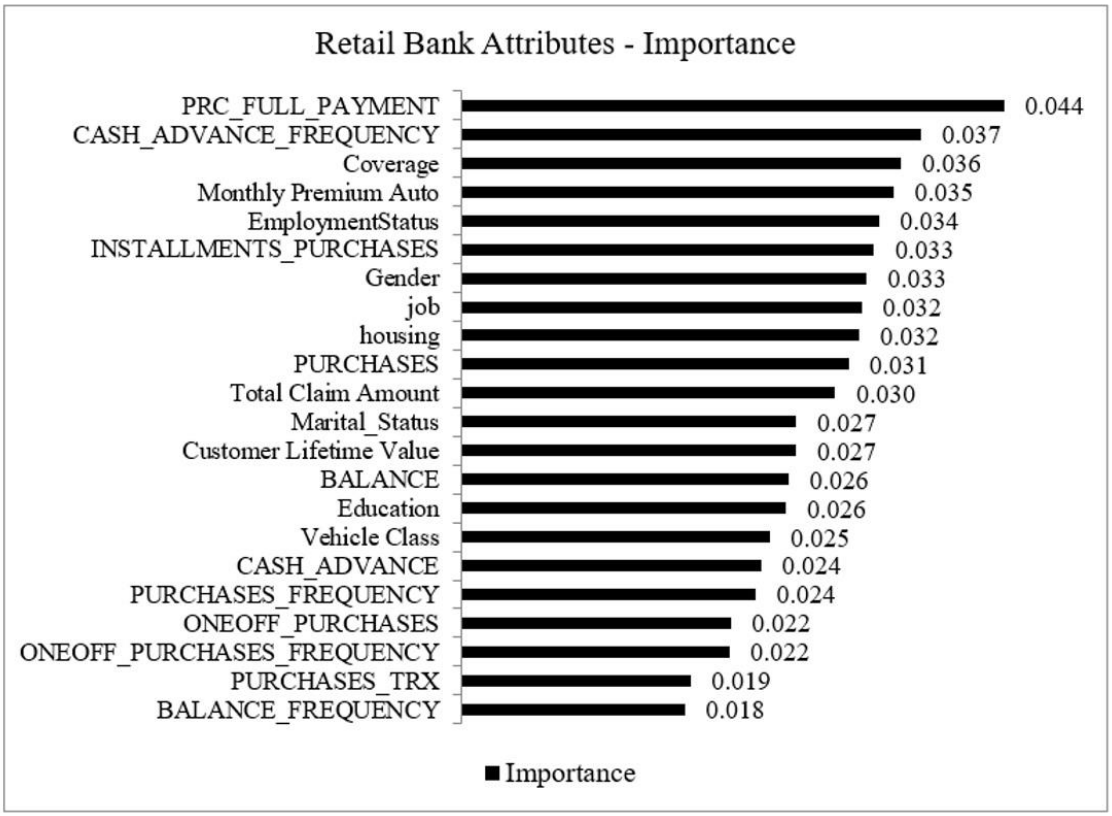

Figure 5. Importance of retail banking attributes.

To identify the best model for predicting customer preferences across all six customer segments, data exploration and due diligence were used to establish the same set of independent and dependent variables. Following a standard development and validation process, multiple machine learning algorithms were tested. Ultimately, XGBoost and Gradient Boosting emerged as the top performers based on model diagnostics. Additionally, a random forest model was employed to understand the relative importance of controllable attributes, allowing businesses to take the targeted actions for optimization. Analysis of all six customer segments identified the most accurate models for product category recommendations in segments 3 and 5 (**Table 7**). These segments can leverage their respective top-performing models for product suggestions. Additionally, the top 10 most influential retail banking attributes were pinpointed through the models (**Table 8** and **Figure 5**), providing valuable insights for developing targeted lead generation campaigns and customer acquisition strategies.

6. Results and Discussions

Retail banks utilize portfolio segmentation to cater to the specific needs and preferences of different customer segments, thereby providing personalized experiences, increasing satisfaction, and retaining customers. K-means clustering is a well-established methodology for retail bank portfolio segmentation. Li et al. (2021) introduced an enhanced customer segmentation method using an improved K-means algorithm and adaptive particle swarm optimization (PSO). They utilized an adaptive learning PSO (ALPSO) to optimize the K-means cluster centers, referred to as KM-ALPSO, thus reducing reliance on initial cluster centers. Factor analysis was applied to extract numerical variables from a mixed dataset of actual group-customer consumption data. The proposed dissimilarity measurement method was employed for clustering the mixed data, showcasing superior performance compared to existing models and demonstrating the effectiveness of KM-ALPSO for customer segmentation. Durga et al. (2023) performed extensive study and identified that customer segmentation analysis involved categorizing a diverse customer base into distinct segments with unique characteristics and preferences of their own. This enables businesses to personalize their marketing efforts and enhance sales performance. Clustering algorithms play a key role in identifying these segments, providing insights for targeted strategies and resource optimization. Ultimately, data-driven customer segmentation serves as a strategic tool for businesses aiming to thrive in competitive markets and improve sales outcomes. Turkmen (2022) has compared customer segmentation techniques using online retail data, including machine learning models and the RFM approach (Recency, Frequency, Monetary Value of a transaction). The study evaluates the insights for marketing strategies, discusses their strengths and weaknesses, and proposes a methodology for selecting the best clustering approach. Additionally, it includes a literature review on artificial intelligence, ML-based clustering models, and examples of customer segmentation in various industries. Gopalakrishnan (2024) examines how customer segmentation in the banking industry can be enhanced by using the K-Means clustering algorithm to group customers based on financial attributes. This approach allows for tailored marketing strategies and personalized offers, ultimately leading to improved customer engagement and profitability. K-means clustering allows banks to identify high-risk segments within their portfolio. This also helps in implementing targeted risk management strategies for those high-risk segments, thereby reducing potential credit losses.

To summarize, segmentation enables retail banks to create targeted marketing campaigns that are better tailored to the characteristics and behaviours of each segment. This leads to more effective marketing efforts, better response rates, and improved cross-selling opportunities. Understanding the needs of different customer segments through portfolio segmentation can guide product development efforts towards creating innovative offerings that address specific market demands. Effective portfolio segmentation using K-means clustering provides retail banks with a competitive edge by enabling them to better understand their customers' behaviours, preferences, and profitability drivers. Overall, portfolio segmentation through K-

means clustering empowers retail banks with valuable insights that drive strategic decision-making across various aspects of the business, ultimately leading to improved customer satisfaction, risk management, marketing effectiveness, product innovation, and competitive positioning in the market.

Upon analysing the cluster features outlined (**Table 3**), several insights into the portfolio emerge. Cluster-1 is characterized by an average income that is notably lower than that of the other segments, yet it has the second-highest average payment across the six portfolio segments. Cluster-2 stands out for having the lowest average payment and credit limit among the segments, although its average income does not rank as the lowest. Meanwhile, Cluster-5 distinguishes itself with the highest purchase frequency of all segments. Coupled with the highest average credit limit, this suggests that Cluster-5 customers exhibit a propensity for higher spending. The model succinctly conveys the key financial behaviours and characteristics of the different clusters within the portfolio.

Bakhtiari et al. (2013) found that, despite their popularity, affinity card programs do not necessarily lead to higher profitability compared to non-affinity card holders. Their study indicated that sports-based and alumni-based programs are less profitable; however, affinity card customers present lower risk and contribute to reducing the average risk of the customer portfolio. In a similar vein, Dinh and Kleimeier (2007) proposed a credit scoring model for retail loans in Vietnam, focusing on identifying borrower characteristics, calibrating the model to align with bank objectives, and assessing its applicability in transactional versus relationship lending scenarios. Finally, Kamimura et al. (2023) conducted a literature review on optimization methods employed in Credit Scoring Models (CSMs). They found that these models are typically developed using Financial Analysis, Machine Learning, Statistical Techniques, Operational Research and Data Mining Algorithms. The findings of the study have practical implications for the application of CSMs. However, it is important to address legal and ethical considerations in the application of optimization methods to CSMs.

In the context of the present study on data monetization, the purpose is to develop a new customer segmentation strategy based on transactional and behavioural data that aims to generate leads and acquire customers for both the bank and the e-commerce platforms. By analysing a customer's profile, the system can assign them to a particular segment. For the bank, this triggers a scoring engine to assess their propensity for e-commerce products and recommend relevant items. Conversely, during an e-commerce transaction, the customer's segment triggers the scoring engine to evaluate their credit card interest, potentially leading to a targeted product offer from the bank. This approach creates a two-way path for lead generation and customer acquisition.

The first model, which is developed based on e-commerce attributes, will be used to predict the propensity of purchase using a credit card during any banking transaction for purchase. A credit card purchase propensity model, built on customer behavior on the e-commerce platform, predicts how likely someone is to use a credit card during a bank transaction. This model achieved the best results for portfolio segment 2, with an impressive 88% accuracy, making it ideal for implementation by the e-commerce company. Importantly, the model also identifies key factors influencing credit card usage, such as review scores, product pricing, payment options, descriptions, and photos. These controllable variables give the e-commerce company the power to optimize these aspects and potentially boost conversions. This not only fosters lead generation through collaboration with the bank but also increases sales and revenue.

The second model, which is developed based on retail banking attributes, will be used to predict product category which customers usually purchase while performing any e-commerce transaction. This model achieved optimal results for portfolio segment 3, making it ideal for the bank to implement. The model also

identifies key banking features influencing purchase behavior, such as account balance activity, transaction history, loan types, and vehicle ownership. These insights can be leveraged for lead generation and customer acquisition within the e-commerce industry, fostering cross-selling opportunities for other banking products and services. Furthermore, by monitoring these factors, the bank can enhance customer satisfaction and potentially reduce churn. Ultimately, this model and its identified features empower the bank to differentiate itself in the market.

7. Managerial Implications

This cross-industry collaboration framework successfully combined data from both the bank and the e-commerce platforms. By culminating in the development of targeted propensity models for each industry, the framework creates a win-win situation. It ensures secure data exchange while simultaneously providing valuable leads to both the bank and the e-commerce company, ultimately boosting their respective businesses.

E-commerce businesses can leverage this framework to identify potential customers during transactions and channel them towards the partnered retail bank's credit card program. To maximize conversions, the bank should develop customized offers and discounts tailored to these leads. To further enhance lead nurturing, the bank can share campaign strategies informed by spending patterns of their existing credit card customer base. This collaboration creates a favourable situation for both the entities, with e-commerce companies gaining qualified leads and banks acquiring new credit card customers.

For retail bank implementation, the second model identifies potential customers interested in essential or luxury goods. The bank can then encourage purchases in these categories by directing these leads to the partnered e-commerce platform with enticing offers. To create these targeted promotions, the e-commerce company should analyse their existing customer base and share their campaign strategies with the bank in advance. This collaboration fosters mutual benefit, with the bank driving sales and the e-commerce platform acquiring new customers. The benefit of such data exchange programme is to get more successful target campaigns and to ensure higher return on investment.

8. Conclusions and Future Steps

This study explores the concept of lead generation through collaboration between retail banks and e-commerce platforms, a relatively under-researched area compared to traditional B2B collaborations. An extensive literature review was conducted to understand the existing practices and identify gaps. This paper proposes a novel framework that leverages propensity models to identify potential customers for both industries. This framework offers a unique approach towards customers' experience enhancement. By analyzing customer data, the model tailor’s product recommendations and promotions to better suit individual needs. This translates to increased customer satisfaction for both retail bank customers (receiving relevant e-commerce offers) and e-commerce customers (benefitting from targeted banking promotions). Ultimately, the framework establishes a win-win strategy for both businesses, generating valuable leads and fostering customer acquisitions.

The study further validates the effectiveness of the framework through the development of two highly accurate propensity models, one for the retail bank and one for the e-commerce platform. Additionally, the application of these models within distinct customer segments, identified using unsupervised machine learning algorithms, ensures precise targeting and a highly effective approach.

The framework developed in this work will have advantages and challenges in order to be implemented in the context of banking industry. However, following this new integrated framework and conducting the

study by using the methodology proposed, will hopefully ensure significant improvement in customer experience and customer management. The proposed framework is expected to be competent enough for application by banking industry. The application of the framework for improvement in customer experience is found to be quite new, as not many published case studies are found. It may be mentioned that a straight-jacketed approach may not be effective in these situations and portfolio segmentation using composite variables (involving retail banks & collaborative industries) is essential for appropriate campaigns & offerings.

This study, while proposing a novel framework for customer experience improvement through bank-e-commerce collaboration, acknowledges limitations. Data access, privacy concerns, and model generalizability across different customer bases are potential hurdles. The framework's reliance on customer acceptance of targeted promotions and ethical considerations around data monetization necessitate further exploration. Additionally, the study's focus on a single industry collaboration and evaluation at a single point in time limits its generalizability and understanding of long-term impact. Finally, potential bias within machine learning algorithms used for segmentation and modeling requires mitigation strategies. Addressing these limitations will strengthen the practical applications of the framework and pave the way for future research.

Finally, this study lays the groundwork for a comprehensive customer experience framework. Here's how we envision its future development:

a) Quantifying Customer Experience: The research would aim to develop a metric that captures both customer satisfaction and dissatisfaction. This will allow to objectively measure the impact of various initiatives on overall customer experience.

b) Leveraging Digital Footprints: The framework can be expanded to incorporate digital footprints, which encompass customer interactions across online channels. By analysing these footprints, one can assess the influence of digital technologies on customer experience and identify areas for improvement.

c) Customer 360° Benchmarking: The final stage involves creating a benchmark that underscores the importance of a holistic customer view (Customer 360°). Here, the enhanced framework, incorporating both digital footprints and cross-industry collaboration data, will be used to evaluate the effectiveness of this approach in driving positive customer experiences. This comprehensive customer view will be the focus of our future research endeavours.

Conflict of Interest
The authors confirm that there is no conflict of interest to declare for this publication.

AI Disclosure
The author(s) declare that no assistance is taken from generative AI to write this article.

Acknowledgments
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editor and anonymous reviewers for their comments that help improve the quality of this work.

ANNEXURE-1

Details of the data sources used in the study.

Data source	Source no.	Dataset details
Retail Banking	1.	Bank-marketing-dataset - Demographic and Campaign details
	2.	Campaign Dataset - Demographic and Campaign details
	3.	Bank-marketing - Demographic, Campaign details and marketing specific variables
	4.	Bank-Marketing-Campaign-dataset - Demographic, Campaign details and marketing specific variables
	5.	Portuguese-Bank-Marketing-dataset - Demographic, Campaign details and marketing specific variables
	6.	Bank-Marketing-dataset - Demographic, Campaign details and marketing specific variables
	7.	Direct-Marketing - Nature of Data: Demographic
	8.	Banking-dataset-Marketing-Targets - Demographic, Campaign details and marketing specific variables
	9.	Credit Card - Credit Card specific details along with customer details
	10.	Email-Campaign-Management - Campaign specific details
	11.	Banking-dataset-marketing-targets (2) - Demographic, Campaign details and marketing specific variables
	12.	Customer-Retention- Campaign for customer retention
	13.	Bank Personal Loan - Customer Credit worthiness specific variables
Cross Industry collaborator: E-commerce Data set	14.	Brazilian-Ecommerce - Complete e-commerce data: (Sub data sets) - Seller - Orders details - Payment details - Customers details - Order review - Geolocation - Order items - Product details - Product category
Cross Industry collaborator: Auto Insurance Data set	15.	Auto Insurance - IBM - Watson - Marketing - Customer - Value – data

ANNEXURE-2

Distribution of all attributes across business functions.

Distribution of 93 attributes across Retail Bank, E-commerce, Credit Card and Auto Insurance				
E-Commerce (32 - Attributes)	Auto Insurance (18 - Attributes)	Credit Card Info (17 - Attributes)	Retail Bank Campaign Data (15 - Attributes)	Retail Bank demography, account and loan (11 - Attributes)
order_id	Customer Lifetime Value	BALANCE	contact	customer_id
product_id	Response	BALANCE_FREQUENCY	month	age
seller_id	Coverage	PURCHASES	duration	job
shipping_limit_date	Effective To Date	ONEOFF_PURCHASES	campaign	marital
price	Gender	INSTALLMENTS_PURCHASES	pdays	education
freight_value	Income	CASH_ADVANCE	previous	default
review_score	Monthly Premium Auto	PURCHASES_FREQUENCY	poutcome	balance
review_comment_title	Months Since Last Claim	ONEOFF_PURCHASES_FREQUENCY	y	housing
review_comment_message	Months Since Policy Inception	PURCHASES_INSTALLMENTS_FREQUENCY	day_of_week	loan
review_creation_date	Number of Open Complaints	CASH_ADVANCE_FREQUENCY	emp.var.rate	deposit
review_answer_timestamp	Number of Policies	CASH_ADVANCE_TRX	cons.price.idx	id
order_status	Policy Type	PURCHASES_TRX	cons.conf.idx	
order_purchase_timestamp	Policy	CREDIT_LIMIT	euribor3m	
order_approved_at	Renew Offer Type	PAYMENTS	nr.employed	
order_delivered_carrier_date	Sales Channel	MINIMUM_PAYMENTS	day	
order_delivered_customer_date	Total Claim Amount	PRC_FULL_PAYMENT	subscribed	
order_estimated_delivery_date	Vehicle Class	TENURE		
product_category_name	Vehicle Size			
product_name_lenght				
product_description_lenght				
product_photos_qty				
product_weight_g				
product_length_cm				
product_height_cm				
product_width_cm				
product_category_name_english				
customer_zip_code_prefix				
customer_city				
customer_state				
seller_zip_code_prefix				
seller_city				
seller_state				

Distribution of shortlisted attributes across business functions.

50 Shortlisted Attributes			
E-Commerce (18 attributes)	Credit Card Info (14 attributes)	Retail Bank Demograpghy & Loan info (11 attributes)	Auto Insurance (7 attributes)
order_id	BALANCE_FREQUENCY	age	Coverage
product_id	PURCHASES	job	Total Claim Amount
price	ONEOFF_PURCHASES	Marital_Status	Monthly Premium Auto
freight_value	INSTALLMENTS_PURCHASES	Education	Customer Lifetime Value
product_category_name	CASH_ADVANCE	EmploymentStatus	Gender
product_name_lenght	PURCHASES_FREQUENCY	Location Code	Income
product_description_lenght	ONEOFF_PURCHASES_FREQUENCY	default	Vehicle Class
product_photos_qty	PURCHASES_INSTALLMENTS_FREQUENCY	BALANCE	
review_id	CASH_ADVANCE_FREQUENCY	housing	
review_score	PURCHASES_TRX	loan	
review_comment_title	CREDIT_LIMIT	Pseudo_id	
review_comment_message	PAYMENTS		
order_status	MINIMUM_PAYMENTS		
expected_delivery_duration	TENURE		
ecom_payment_sequential			
ecom_payment_type			
ecom_payment_installments			
ecom_payment_value			

References

Bader, K. (2013). How to benefit from cross-industry innovation? A best practice case. *International Journal of Innovation Management*, *17*(6), 1340018. https://doi.org/10.1142/s1363919613400185.

Bakhtiari, A., Murthi, B.P.S., & Steffes, E. (2013). Evaluating the effect of affinity card programs on customer profitability using propensity score matching. *Journal of Interactive Marketing*, *27*(2), 83-97. https://doi.org/10.1016/j.intmar.2012.10.003.

Beck, R., Wigand, R.T., & König, W. (2005). The diffusion and efficient use of electronic commerce among small and medium-sized enterprises: an international three-industry survey. *Electronic Markets*, *15*(1), 38-52.

Bellingkrodt, S., & Wallenburg, C.M. (2015). The role of customer relations for innovativeness and customer satisfaction: A comparison of service industries. *The International Journal of Logistics Management*, *26*(2), 254-274. https://doi.org/10.1108/ijlm-06-2012-0038.

Chan, F.T.S., Chong, A.Y.L., & Zhou, L. (2012). An empirical investigation of factors affecting e-collaboration diffusion in SMEs. *International Journal of Production Economics*, *138*(2), 329-344. https://doi.org/10.1016/j.ijpe.2012.04.004.

Dinh, T.H.T., & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis*, *16*(5), 471-495. https://doi.org/10.1016/j.irfa.2007.06.001.

Durga, P.S., Paulson, J.A., & Srinivasareddy, M. (2023). Customer segmentation analysis for improving sales using clustering. *International Journal of Science and Research Archive*, *09*(02), 708-715. https://doi.org/10.30574/ijsra.2023.9.2.0663.

Eid, R., Trueman, M., & Ahmed, A.M. (2002). A cross-industry review of B2B critical success factors. *Internet Research*, *12*(2), 110-123. https://doi.org/10.1108/10662240210422495.

Francis, M. (2024). Data monetization: utilizing data as an asset to produce new revenue for modern businesses. *https://www.researchgate.net/publication/378437368*. https://doi.org/10.13140/rg.2.2.28202.59846.

Gilbert, G.R., & Veloutsou, C. (2006). A cross-industry comparison of customer satisfaction. *Journal of Services Marketing*, *20*(5), 298-308. https://doi.org/10.1108/08876040610679918.

Gopalakrishnan, K. (2024). Customer segmentation using K-means clustering for targeted marketing in banking. *International Journal of Artificial Intelligence & Machine Learning*, *3*(2), 89-94. https://doi.org/10.5281/zenodo.13627403.

Hanafizadeh, P., & Nik, M.R.H. (2020). Configuration of data monetization: a review of literature with thematic analysis. *Global Journal of Flexible Systems Management*, *21*(1), 17-34. https://doi.org/10.1007/s40171-019-00228-3.

Hånell, S.M., Nordman, E.R., Tolstoy, D., & Özbek, N. (2020). "It's a new game out there": e-commerce in internationalising retail SMEs. *International Marketing Review*, *37*(3), 515-531. https://doi.org/10.1108/imr-03-2018-0107.

Jamali, D. (2007). A study of customer satisfaction in the context of a public-private partnership. *International Journal of Quality & Reliability Management*, *24*(4), 370-385. https://doi.org/10.1108/02656710710740545.

Jupp, J.R. (2016). Cross-industry learning: a comparative study of product lifecycle management and building information modelling. *International Journal of Product Lifecycle Management*, *9*(3), 258-284. https://doi.org/10.1504/ijplm.2016.080502.

Kamimura, E.S., Pinto, A.R.F., & Nagano, M.S. (2023). A recent review on optimization methods applied to credit scoring models. *Journal of Economics, Finance and Administrative Science*, *28*(56), 352-371. https://doi.org/10.1108/jefas-09-2021-0193.

Kulhanek, O. (2019). *The five rules of data monetization*. KPMG. https://kpmg.com/xx/en/home/insights/2019/04/the-five-rules-of-data-monetization.print.html.

Kurnia, S., Choudrie, J., Mahbubur, R.M., & Alzougool, B. (2015). E-commerce technology adoption: a Malaysian grocery SME retail sector study. *Journal of Business Research*, *68*(9), 1906-1918. https://doi.org/10.1016/j.jbusres.2014.12.010.

Lee, S.C., Pak, B.Y., & Lee, H.G. (2003). Business value of B2B electronic commerce: the critical role of inter-firm collaboration. *Electronic Commerce Research and Applications*, *2*(4), 350-361. https://doi.org/10.1016/s1567-4223(03)00003-6.

Lemon, K.N., & Verhoef, P.C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, *80*(6), 69-96. https://doi.org/10.1509/jm.15.0420.

Li, Y., Chu, X., Tian, D., Feng, J., & Mu, W. (2021). Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm. *Applied Soft Computing*, *113*(B), 107924. https://doi.org/10.1016/j.asoc.2021.107924.

Luo, Y., & Zahra, S.A. (2023). Industry 4.0 in international business research. *Journal of International Business Studies*, *54*(3), 403-417. https://doi.org/10.1057/s41267-022-00577-9.

Ofulue, J., & Benyoucef, M. (2024). Data monetization: insights from a technology-enabled literature review and research agenda. *Management Review Quarterly*, *74*(2), 521-565. https://doi.org/10.1007/s11301-022-00309-1.

Olaleye, S.A., & Adusei, A.G. (2024). The new reality of data economy and productization: a conceptual paper. *Finnish Business Review*. http://urn.fi/urn:nbn:fi:jamk-issn-2341-9938-83.

Ondrus, J., & Pigneur, Y. (2007). Cross-industry preferences for development of mobile payments in Switzerland. *Electronic Markets*, *17*(2), 142-152. https://doi.org/10.1080/10196780701296386.

Padmavathy, C., Balaji, M.S., & Sivakumar, V.J. (2012). Measuring effectiveness of customer relationship management in Indian retail banks. *International Journal of Bank Marketing*, *30*(4), 246-266. https://doi.org/10.1108/02652321211236888.

Raja, J.Z., Bourne, D., Goffin, K., Cakkol, M., & Martinez, V. (2013). Achieving customer satisfaction through integrated products and services: an exploratory study. *Journal of Product Innovation Management*, *30*(6), 1128-1144. https://doi.org/10.1111/jpim.12050.

Rao, B. (1999). The internet and the revolution in distribution: a cross-industry examination. *Technology in Society*, *21*(3), 287-306. https://doi.org/10.1016/s0160-791x(99)00018-4.

Ritala, P., Keränen, J., Fishburn, J., & Ruokonen, M. (2024). Selling and monetizing data in B2B markets: four data-driven value propositions. *Technovation*, *130*, 102935. https://doi.org/10.1016/j.technovation.2023.102935.

Saragih, L.R., Dachyar, M., & Zagloel, T.Y.M. (2021). Implementation of telecommunications cross-industry collaboration through agile project management. *Heliyon*, *7*(5), e07013. https://doi.org/10.1016/j.heliyon.2021.e07013.

Sarkar, S., & Sarkar, P. (2019). Data monetization strategies for new age digital communication providers in India. *Telecom Business Review*, *12*(1), 23-34. http://publishingindia.com/tbr/.

Tarafdar, M., & Vaidya, S.D. (2006). Challenges in the adoption of e-commerce technologies in India: the role of organizational factors. *International Journal of Information Management*, *26*(6), 428-441. https://doi.org/10.1016/j.ijinfomgt.2006.08.001.

Turkmen, B. (2022). Customer segmentation with machine learning for online retail industry. *The European Journal of Social & Behavioural Sciences*, *31*(2), 111-136. https://doi.org/10.15405/ejsbs.316.

Zhu, K. (2014). The complementarity of information technology infrastructure and e-commerce capability: a resource-based assessment of their business value. *Journal of Management Information Systems*, *21*(1), 167-202. https://doi.org/10.1080/07421222.2004.11045794.

Zott, C., Amit, R., & Donlevy, J. (2000). Strategies for value creation in e-commerce: best practice in Europe. *European Management Journal*, *18*(5), 463-475. https://doi.org/10.1016/s0263-2373(00)00036-0.

Zwass, V. (2003). Electronic commerce and organizational innovation: aspects and opportunities. *International Journal of Electronic Commerce*, *7*(3), 7-37. https://doi.org/10.108/10864415.2003.11044273.

Original content of this work is copyright © Ram Arti Publishers. Uses under the Creative Commons Attribution 4.0 International (CC BY 4.0) license at https://creativecommons.org/licenses/by/4.0/

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.