ATENEO DE MANILA UNIVERSITY

An Integrated Approach to Next Best Offer Modeling for Financial Products: A Multi-Model Architecture

A THESIS SUBMITTED TO
THE GRADUATE FACULTY OF
THE SCHOOL OF SCIENCE AND ENGINEERING
IN CANDIDACY FOR THE DEGREE OF
MASTER IN
DATA SCIENCE

DEPARTMENT OF INFORMATION SYSTEMS
AND COMPUTER SCIENCE

BY
BENJAMIN LOUIS L. ANG
DANE LAUREN N. ROSARIO

QUEZON CITY, PHILIPPINES
APRIL 2025

The THESIS entitled:

An Integrated Approach to Next Best Offer Modeling for Financial Products: A Multi-Model Architecture

submitted by **Benjamin Louis L. Ang** and **Dane Lauren N. Rosario** has been examined and is recommended for **Oral Defense**.

PATRICIA ANGELA R. ABU, Ph.D.
Chair

CLARK KENDRICK C. GO, Ph.D.
Adviser

RAPHAEL A. GUERRERO, Ph.D.
Dean
School of Science and Engineering

The Faculty of the Department of Information Systems and Computer Science, School of Science and Engineering, Ateneo de Manila University ACCEPTS THE THESIS entitled:

An Integrated Approach to Next Best Offer Modeling for Financial Products: A Multi-Model Architecture

submitted by **Benjamin Louis L. Ang** and **Dane Lauren N. Rosario** in partial fulfillment of the requirements for the degree of Master in Data Science.

JOHN PAUL C. VERGARA, Ph.D.
Member

VICTOR ANDREW ANTONIO, Ph.D.
Member

TERRENCE ERARD D. TEH, MS
Member

CLARK KENDRICK C. GO, Ph.D.
Adviser

RAPHAEL A. GUERRERO, Ph.D.
Dean
School of Science and Engineering

Grade:
Date: April 1, 2025

ABSTRACT

Personalized engagement is vital in banking, but recommendations are often fragmented and product-focused. This study designed, implemented, and evaluated a centralized next best offer (NBO) system for automated, customer-centric offers. Using customer demographic and product ownership data, a hybrid engine combined customer similarity, product similarity, and XGBoost propensity modeling. Effectiveness was rigorously tested via leave-one-out and temporal validation (metrics: hit Rate, HR@K, MRR). Results showed the propensity model generally achieved the best performance across most product categories. However, other models demonstrated strengths for specific products or metrics (e.g., customer similarity for investment products), highlighting model complementarity. An ensemble using greedy weight matrix optimization targeting Hit Rate@3 was also implemented. While demonstrating the potential of combining model strengths, the ensemble provided only marginal improvements over the top-performing individual model. Temporal validation indicated predicting sequential acquisitions remains challenging. The study provides a practical NBO framework and insights into the relative performance of different recommendation techniques in banking.

ACKNOWLEDGMENTS

First and foremost, we would like to express our sincere gratitude to our thesis adviser, Dr. Clark Kendrick C. Go, for sharing his invaluable guidance and wisdom throughout this research. His support was instrumental in the completion of this thesis.

Secondly, we extend our profound thanks to our partner bank. Their team went above and beyond to ensure we had all the necessary guidance and support.

Thirdly, we would like to thank our panelists, Dr. John Paul C. Vergara, Dr. Victor Andrew Antonio, and Mr. Terrence Erard D. Teh, for dedicating their time and effort to provide constructive criticism and valuable suggestions that significantly improved our work.

Finally, we are grateful to our family and friends for providing us with unwavering support as we pursued this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Tables Page

CHAPTER I

INTRODUCTION

1.1 Introduction

In today's banking industry, personalized customer engagement is crucial for building lasting relationships and meeting evolving customer expectations [20]. Customers seek relevant and timely recommendations that fit their financial needs, and banks are increasingly using data-driven approaches to deliver these personalized offers [1]. To address this challenge, this paper proposes the implementation of a next best offer (NBO) system. An NBO system is a powerful tool that enables banks to recommend products based on customer data and predicted preferences across all interaction channels.

One of the leading banks in the Philippines currently manages product recommendations within individual departments, resulting in a fragmented and inconsistent experience for customers. For brevity, we will refer to this institution as "the Bank" throughout this study. This departmentalized approach is neither automated nor data-driven, and relies instead on subjective judgment. Consequently, recommendations tend to be department-centric, lacking a holistic view of each customer's preferences and behaviors. For example, a customer with a retail savings account might be overlooked for offerings like index funds, ETFs, or even higher-yield time deposits managed by the investment division, despite their profile indicating potential suitability based on their demographic and financial characteristics.

To address these limitations, the Bank aims to implement a centralized, automated NBO system that will unify recommendations across all departments and channels. This transformation is intended to shift the Bank's approach from a product-focused model to a customer-centric one, enhancing the consistency and relevance of offers. By leveraging customer demographic data and product ownership history, the new NBO system will enable the Bank to tailor recommendations to each customer's profile, whether they engage with the Bank in person or through digital channels. This shift not only promises a more seamless and personalized experience for customers but also positions the Bank to increase engagement and strengthen customer loyalty through data-driven and precise recommendations.

This study builds on established frameworks for NBO systems in banking, particularly Ziad El Abbass' [1] work on predictive modeling for personalized offers. The proposed NBO architecture will utilize advanced analytics and machine learning, including propensity scoring and similarity matrices, to create a data-driven recommendation engine that optimizes product suggestions.

By automating and centralizing the recommendation process, this NBO system will help the Bank improve customer engagement, increase conversion rates, and provide a seamless, personalized experience across all touchpoints. Beyond its benefits to the Bank, this project contributes valuable insights to the banking industry, offering a model for other financial institutions aiming to modernize their customer engagement strategies through data-driven recommendations.

1.1.1 Research Questions

This study addresses the overarching objective of designing and implementing a centralized next best offer (NBO) system that unifies product recommendations across departments while improving consistency and relevance in customer interactions. To achieve this goal, the research seeks to answer the following specific questions:

1. How can customer demographic information and product ownership history be effectively leveraged within a hybrid recommendation framework to generate personalized and actionable product suggestions?
2. What combination of recommendation techniques—content-based filtering, collaborative filtering, and propensity modeling—provides the most effective framework for product recommendations in the banking context?
3. What validation methodologies and performance metrics most accurately measure the effectiveness of a next best offer system in a banking environment, and how do these metrics translate to potential business outcomes?

1.1.2 Objectives

The primary objective of this study is to design and implement a centralized next best offer (NBO) system for the Bank that unifies product recommendations across all departments and customer touchpoints. This system aims to replace the current fragmented approach with a cohesive recommendation strategy.

Specifically, the study will implement a recommendation engine which is the core of the NBO system for the Bank. The framework incorporates content-based filtering to analyze product characteristics and match them with customer preferences. It also employs collaborative filtering to identify recommendations based on similarity patterns between customers, following a "customers like you also bought" approach. Additionally, the system utilizes propensity models to predict the likelihood of a customer accepting specific products based on their demographic and behavioral data.

This research will utilize the Bank's customer data, which contains both product ownership information and customer demographic characteristics. By analyzing these data points, the study aims to discover meaningful patterns in customer behavior and product selection that can inform more relevant offerings.

A key objective is to evaluate the effectiveness of the NBO system through rigorous validation methods. The study will employ a recommendation validator framework to measure performance metrics such as hit rate, precision, recall, and mean reciprocal rank. This validation approach will provide quantitative evidence of the system's ability to deliver relevant recommendations across different customer segments.

Through these objectives, the study seeks to create a practical, data-driven NBO system that enhances customer experience while providing the Bank with a competitive advantage in customer engagement and product adoption.

1.1.3 Scope and Limitations of the Study

This study aims to develop a recommendation engine as the core of an NBO system tailored to the Bank's business banking segment, leveraging existing customer data to automate and personalize recommendations across a diverse range of product categories. These categories include corporate and retail credit cards, loans, investments, trade services, insurance, cash management solutions, various deposit and savings accounts, foreign exchange services, and remittance products.

To generate accurate and relevant product suggestions, the study will focus on the design and implementation of a hybrid recommendation framework that combines content-based filtering, collaborative filtering, and propensity modeling. The recommendation engine will analyze customer attributes such as demographic information, product ownership patterns, and digital engagement behavior to decide relevant product recommendations.

The scope of the study is limited to the data provided by the Bank, which consists mainly of synthetic data due to privacy restrictions. Although synthetic data mimic real-world customer behavior, it may not fully capture the nuances of actual customer interactions, potentially impacting the accuracy and personalization of recommendations. In addition, the system's personalization will be based solely on existing customer attributes, such as demographic information, product ownership, and digital engagement patterns, without further data collection. This may limit the system's ability to address less tangible customer needs or preferences.

CHAPTER II

REVIEW OF RELATED LITERATURE

The financial system in the Philippines is characterized by a diverse range of institutions, including universal, commercial, and digital banks, each offering various products and services. This sector undergoes continuous transformation, driven by the convergence of economic, demographic, and technological shifts [11]. The Philippines is experiencing one of the fastest population growth rates globally, with the bankable population expected to grow from 65 million in 2022 to 85 million by 2030 [11]. This growth is largely driven by a young and highly digital consumer demographic which fuels an increasing demand for innovative services that cater to their tech-savvy preferences. To meet this rising demand, banks must embrace data-driven approaches to enhance their service offerings. Leveraging customer data through advanced analytics allows institutions to create personalized financial solutions tailored to individual needs, preferences, and behaviors.

One such approach is the next best action (NBA) methodology, which utilizes predictive models to recommend optimal actions, enhancing personalization and decision-making. Key methodologies under NBA include reinforcement learning, which optimizes decision-making through iterative feedback and recommender systems, which provide tailored product suggestions. Within the broader NBA framework, the next best offer (NBO) represents a focused subset that specifically targets marketing and sales contexts. NBO aims to recommend the most relevant product or service by scoring potential offers across predictive models and selecting the most relevant products for the customer. While NBA focuses on optimizing the entire customer experience, NBO narrows its scope to delivering personalized offers that align with customer needs.

This section reviews the application of these methodologies, primarily in non-financial domains, and explores their potential in the Philippine banking sector, where they can enhance customer satisfaction and improve business outcomes.

2.1 Financial System in the Philippines

Banks are financial institutions authorized to accept deposits, such as savings and checking accounts, and provide loans. In addition, they also offer a range of supplementary services, including individual retirement accounts (IRAs), certificates of deposit (CDs), currency exchange, personal loans, and secure storage through safe deposit boxes. These institutions can be broadly classified into retail banks, which cater to individual customers, commercial or corporate banks, which serve businesses, and investment banks, which focus on financial markets and advisory services [16].

Specifically, the Philippine financial system comprises a diverse array of institutions that collectively support the country's economic and financial activities. This system is broadly categorized into banks and non-bank financial institutions (NBFIs), each playing a unique role in the financial ecosystem [4].

2.1.1 Banks

Banks are the cornerstone of the financial system, serving as the primary institutions for deposit-taking and credit provision. The Philippine banking sector is further segmented into the following categories [3]:

- **Universal banks**: These institutions perform both commercial banking activities and additional functions including powers of investment houses and full equity investment in non-allied enterprises.

- **Commercial banks**: Commercial banks provide traditional banking services, including accepting deposits, issuing letters of credit, and extending loans.
- **Thrift banks**: As defined under Republic Act (R.A.) No. 7906, this type of bank serves small and medium enterprises, offering savings and loan facilities [10].
- **Rural banks**: Defined under R.A. No. 7353, this bank caters to rural and agricultural sectors by providing financial services to farmers, fishermen, and other rural communities [9].
- **Cooperative banks**: Cooperative banks provide financial services primarily to cooperatives and their members.
- **Islamic bank**: These banks operate under R.A. No. 6848 and offer financial products that are compliant with Islamic law or Shari'ah principles [8].
- **Digital bank**: A recent innovation, digital banks operate primarily online, offering a wide range of banking services through digital platforms.

2.1.2 Non-Bank Financial Institutions (NBFIs)

NBFIs complement the role of banks by offering specialized financial services, including insurance, investment management, leasing, and other financial activities not covered under traditional banking [4].

With the rise of digital banking platforms and the increasing importance of personalized financial services, there is a growing need for institutions to adopt data-driven approaches that can enhance customer experience and improve service outcomes.

The next section will explore various approaches currently employed in different disciplines which includes reinforcement learning, recommender Systems, and next best offer.

2.2 Reinforcement Learning

Reinforcement learning is a category of machine learning where agents to make decisions by interacting with their environment to get the most optimal results, where optimality varies by case. It is done through a trial-and-error type of learning that reinforces actions that lead towards the goal and punish those that do not [15].

2.2.1 A/B Testing

A/B testing is a form of hypothesis testing where two variants are compared in the field from an end user's point of view [21]. A common application is in product recommendations or marketing messages. However, A/B testing can be time-consuming and resource-intensive. Multi-armed bandits, a type of RL algorithm, offer a more efficient alternative [23].

2.2.2 Multi-armed Bandits

The Multi-armed bandit is an improved version of A/B testing that uses machine learning algorithms so that it can allocate more resources to well-performing assets and less to those that are not [23]. It determines the most profitable outcome through a series of choices. At the beginning of the experiment, when odds and payouts are unknown, the agent must determine which machine to pull, in which order, and how many times [23]. However, there is a need to balance exploration, learning the rewards of each option, and exploitation, going with the asset that gives the most return [23].

2.2.3 Contextual Bandits

Contextual bandits extend the concept of multi-armed bandits by incorporating contextual information into the decision-making process. This means that the recommendation is not solely based on the performance of different options but also takes into account relevant context [23].

In recommender systems, contextual bandits address the exploration-exploitation dilemma by balancing the use of known user preferences with the exploration of new items that may interest the user. For instance, different studies in 2017 proposed large-scale bandit approaches to allow continuous exploration of problems in recommender systems even when prior information is scarce [25] [6]. However, many existing approaches do not consider how users' preferences evolve over time so Bouneffouf proposed an algorithm that dynamically adjusts exploration and exploitation to account for such changes, identifying the most suitable moments for each [5]. This was done through the introduction of the Freshness-Aware Thompson Sampling algorithm, which optimizes the recommendation of fresh content by assessing user risk in different scenarios.

These reinforcement learning techniques, particularly multi-armed bandits and contextual bandits, offer significant advantages in decision-making for personalized recommendations. By efficiently allocating resources to the most promising options, multi-armed bandits can optimize product offerings, improving customer engagement and maximizing returns. The incorporation of contextual information through contextual bandits allows for even more tailored recommendations by factoring in customer behavior and preferences. In the context of the Philippine banking sector, these methodologies can enhance NBO models, enabling banks to offer personalized financial products that better match customer needs, all while minimizing resource usage.

2.3 Recommender Systems

While RL balances the act of exploration and exploitation, the NBA is a customer engagement approach that utilizes a combination of artificial intelligence and real-time interaction data to create hyper-relevant customer experiences. Several recommendation techniques can be employed in NBA systems to provide personalized and relevant suggestions to users or customers [13].

2.3.1 Collaborative Filtering

Collaborative filtering analyzes historical data on similar users or customers to identify patterns and recommend actions that were successful for them. It is classified into two different approaches: (1) memory-based techniques (2) model-based techniques [14].

Memory-based techniques

Memory-based algorithms tackle the collaborative filtering problem by utilizing the entire dataset which includes the users and their ratings that indicates preference [12]. These methods focus on identifying users who share similar preferences with the active user—those for whom predictions are being made—and use their ratings to forecast the active user's preferences.

This method relies on historical data to make recommendations by leveraging the collective behavior of similar users, and there are two approaches.

First is user-based collaborative filtering which identifies groups of users with similar preferences and recommends items that are popular among these groups. This approach involves finding a set of users who liked the same item and determining if a target user will also appreciate it. By analyzing the behavior of similar users, the system can identify potential matches and offer relevant recommendations. On the other hand, item-based collaborative filtering focuses on the similarity between items rather than users. If an item is positively rated by a user, the system recommends other items with similar ratings to that user. This approach views ratings as vectors in an n-dimensional space and calculates similarity based on the angle between these vectors. Additionally, item-based collaborative filtering often incorporates correction mechanisms to improve the precision of similarity analysis [13]. The difference is illustrated below in Figure 2.1.

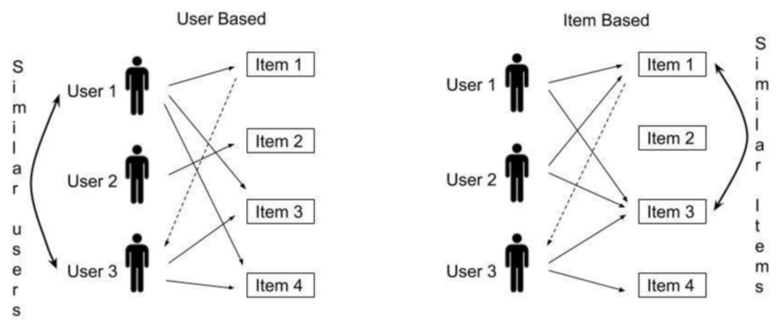


Figure 2.1: User-based and Item-based Collaborative Filtering Diagram from [24]

Model-based techniques

Model-based collaborative filtering techniques employ machine learning algorithms to analyze user rating data and discover underlying patterns. These models aim to improve the accuracy and efficiency of recommendations compared to memory-based approaches.

Unlike memory-based methods, model-based approaches create a model beforehand to compute recommendations. Model-based approaches gained significant traction after Netflix's million-dollar prize competition in 2006. As the competition progressed, more participants turned to these methods with the winning team primarily using model-based matrix factorization techniques [17].

Matrix factorization decomposes the user-item rating matrix into two lower-dimensional matrices, representing latent user and item features. By calculating the dot product of these latent features, the model can predict missing ratings and provide personalized recommendations. Several dimensionality reduction techniques can be applied to build model-based collaborative filtering systems, including principal component analysis (PCA), probabilistic matrix factorization (PMF), and singular value decomposition (SVD) [14].

This was further supported by a comparative study conducted to assess the effectiveness of memory-based and model-based collaborative filtering techniques for e-commerce recommendations in Indonesia [2]. The results indicated that model-based approaches outperformed memory-based methods in terms of recommendation accuracy, computational efficiency, and relevance.

2.3.2 Content-Based Filtering

Content-based filtering utilizes information about the user or customer's preferences and behavior to recommend actions that align with their interests. For example, the system might recommend similar products or related services if a user has expressed interest in a particular product category [19].

2.3.3 Hybrid Filtering

Hybrid filtering combines the advantages of collaborative and content-based filtering to create a more comprehensive recommending system [19]. These systems leverage the strengths of both techniques to provide more personalized and relevant suggestions.

Recommender systems offer promising methodologies for providing personalized product recommendations, yet their practical application in the banking sector faces several challenges. One of the primary obstacles is the difficulty in obtaining accurate and comprehensive data on customer ratings and preferences for bank products. Without sufficient and reliable data, these systems struggle to deliver highly personalized recommendations, which can hinder their effectiveness. Additionally, the banking industry must consider eligibility rules and regulatory requirements before recommending products to customers. Simply offering products without considering these critical factors can result in irrelevant or even harmful recommendations. As a result, while the methodologies explored in this section hold significant potential, their direct implementation in banks may require substantial adjustments, particularly in data collection and compliance with industry regulations.

2.4 Model Validation

Model validation is the performance assessment of a trained model, especially in cases of new or unseen data. It assures the model's performance is as prescribed and for the intended purpose when executed in the real world. A number of different model validation techniques have been established to assess the predictive ability of models. The most common model validation techniques are given in the following subsections [18].

2.4.1 Train/Test Split

With the train/test split method, the dataset is divided into a training and test set. The training set is used to train the model, while the test set acts to evaluate the model performance against unseen data. Typical split ratios for the two are 70-30 or 80-20 where the larger portion is allocated for training. While this approach is straightforward, the evaluation results can vary depending on how the data is partitioned.

2.4.2 k-Fold Cross-Validation

In k-Fold Cross-Validation, the dataset is split into k equally sized subsets (folds). The model is trained and tested k times, each time using a different fold as the test set while the remaining folds are used for training. The final performance score is given from an average of all fold results. This technique reduces the impact of data partitioning and provides a more stable estimate of model performance.

2.4.3 Leave-One-Out Cross-Validation (LOOCV)

LOOCV is a special case of k-Fold Cross-Validation where the number of folds equals the total number of data points. In this method, the model is trained on all data points except one and evaluated on the excluded point. This process repeats for each data point, and the final performance is averaged across all iterations. While LOOCV offers a thorough evaluation, it can be computationally demanding for large datasets.

2.4.4 Leave-One-Group-Out Cross-Validation

This method is employed for data that are already naturally clustered or grouped. The model is trained on one or more groups in each of the iterations, leaving an entire group out for testing. By so doing, the performance of the model is tested against data samples that were not included in the training data, thereby preventing it from overfitting certain patterns of the groups.

2.4.5 Nested Cross-Validation

Nested Cross-Validation is a process of carrying out two levels of cross-validations: the inner loop is responsible for tuning hyperparameters, and the outer loop is responsible for evaluating the performance of a model. More specifically, the inner loop selects the best possible model configuration, and the outer loop examines how well the model generalizes on unseen data. This method gives a reliable estimate on how well the model can perform, although tuning can get rather complicated.

Various model validation techniques provide various strengths and weaknesses that differ depending on the dataset and model. Given that customer behavior patterns and product ownership data in this study are core components of the recommendation process, it is important to choose a validation method that accurately reflects real-world conditions. The particular method of validation that is employed in this research will be elaborated on in the Methodology section.

2.5 Next Best Offer (NBO)

While recommender systems offer great potential for personalized recommendations, their application in banking presents significant challenges, particularly in terms of data availability and regulatory constraints. These obstacles often hinder the ability of traditional recommendation models to provide highly personalized and contextually relevant suggestions. To address these limitations, organizations are increasingly turning to more scalable and customer-centric approaches. An approach suggested by Fabrizi is the implementation of next best offer (NBO) systems [13]. She argued that businesses are striving to become more customer-centric by prioritizing customer needs and preferences throughout product development and marketing processes so a more robust recommender system is needed.

As a result, businesses have turned to data-driven strategies to achieve customer-centricity at scale [13]. Her proposed NBO system leverages customer data and advanced analytics to identify the most relevant and personalized offers for each customer. By providing customers with tailored recommendations, businesses can enhance customer engagement, increase sales, and foster long-term loyalty.

The NBO architecture as seen in Figure 2.2 consists of three interconnected modules.

1. **Information Gathering Module**: This section gathers comprehensive data about the client, including product catalog, client preferences, unowned products, purchase history, reactions to offers, and past transactions.

2. **Recommendation Engine:** The core of the system, which analyzes the collected information using data science and machine learning techniques to identify the most likely product to be purchased by the client.

3. **Interaction Channels Module:** This module manages client communi-

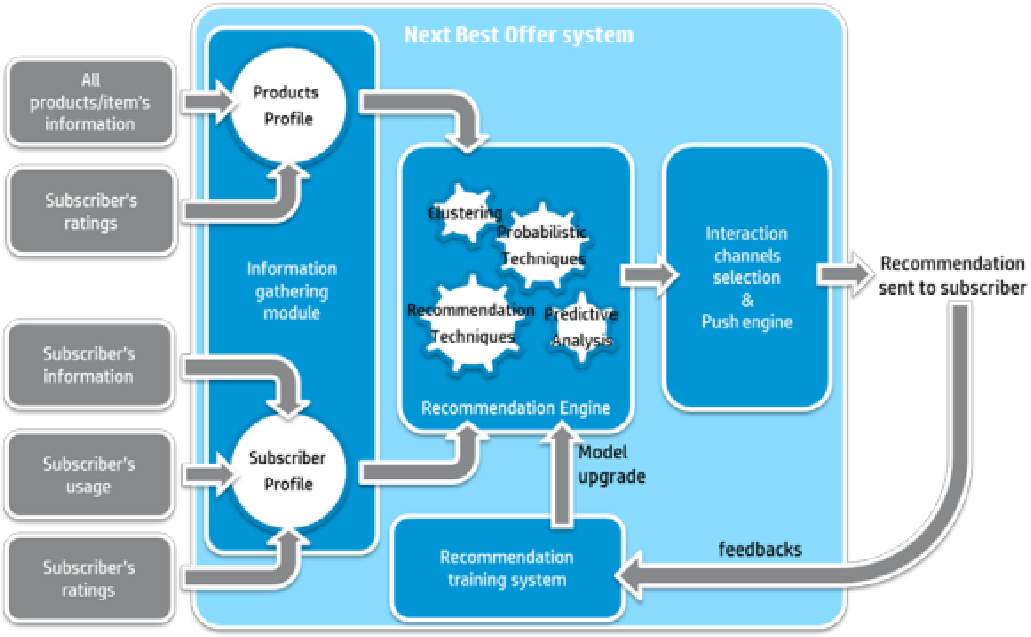


Figure 2.2: NBO Architecture from [13]

cation strategies, taking into account key factors such as preferred communication channels and optimal timing for contact. By effectively managing interaction channels, this module increases the probability of successful offers.

By combining detailed client data, advanced analytics, and personalized communication strategies, this architecture optimizes the NBO system. By effectively leveraging these components, tailored recommendations can be provided that enhance customer satisfaction and drive business growth.

This solution is usually recommended to companies who want to improve their recommendation systems [1]. However, it is challenging for complex businesses like banks to implement it due to various challenges. The bank's wide range of products and services complicates the direct application of an NBO model. The diversity of offerings can make it difficult to accurately categorize and compare products, hindering the ability to effectively recommend suitable options to clients. In addition, the absence of a standardized system for rating products based on client usage poses a significant challenge. Without a consistent framework for evaluating product performance, it is difficult to accurately assess customer satisfaction and preferences, which are crucial for effective NBO recommendations. There is also a limitation on the availability of customer data that can hinder the accuracy of NBO systems as the system may not have sufficient information to assess customer preferences for a wider range of products. Lastly, before making recommendations to clients who may be financially unstable, it is essential to carefully consider risks and ensure that the recommended products align with the client's financial situation. Failure to do so could lead to negative consequences for both the bank and the client [1].

As a response to this limitation, a theoretical roadmap as illustrated in Figure 2.3 was created in 2018 to cater to large banking enterprises in the form of the five proposed pillars for an improved NBO system [1].

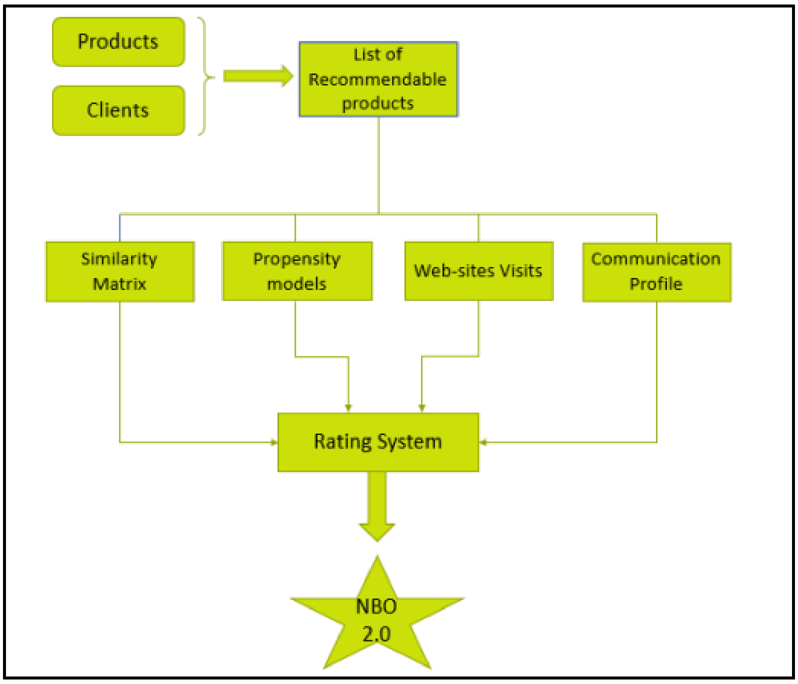


Figure 2.3: Millenium BCP NBO 2.0 Diagram from [1]

1. **List of Recommendable Products** is comprised of a comprehensive list of potential recommendations that are available to the clients. This list is carefully curated based on product categories and client eligibility criteria.

2. **Communication Profile of the Client** involves a detailed analysis of client behavior across various channels. Client behavior is analyzed across various channels and receptivity rates are calculated for each product category. By understanding client preferences, the system can identify optimal communication channels and timing for delivering personalized offers.

3. **Recommendation System** is the third pillar where a two-tiered recommendation system was proposed. The outer recommendation system focuses on recommending product families, while the inner recommendation system focuses on recommending specific items within the selected product family. This approach allows for a more granular and personalized recommendation process.

4. **Website Visits** can get an active behavior that can indicate their interests and preferences by tracking webpage interactions. This information can be used to identify products that clients are actively considering and tailor recommendations accordingly.

5. **Predictive Model** includes the developed predictive models of the bank to analyze client behavior and predict the likelihood of purchasing specific products.

To determine the most suitable product category for each client, a scoring system is implemented. Each pillar contributes a certain number of points to the overall score of a product category which can be arbitrarily chosen by the researchers. The product category with the highest score is considered the most likely to be of interest to the client and is therefore the next best offer.

2.6 Research Gap

The current NBO systems in some Philippine banks face several limitations. This includes a (1) lack of automation, (2) reliance on departmental processes, and (3) a product-centric approach. These systems often involve manually curated recommendations by department staff, based on their perceptions and subjective judgment of customer needs. This process is not automated and is driven by an understanding of the customer that is limited to their immediate interactions with the bank, often relying on the subjective judgement of the employees. As a result, product recommendations are typically centered around the bank's offerings rather than the holistic financial journey of the customer. Moreover, there is also a struggle with scalability and consistency as recommendations are often reactive and not tailored to individual customer preferences or behaviors.

This study seeks to address these gaps by proposing a more automated, customer-centric NBO system that integrates advanced data analytics and machine learning models, enabling the bank to provide more relevant, personalized product recommendations. By bridging these limitations, our approach aims to enhance customer satisfaction and loyalty, ensuring that product offerings are aligned with individual customer needs and financial situations.

CHAPTER III

METHODOLOGY

This study adapts the next best offer (NBO) recommender system architecture proposed by El Abbass [1], focusing specifically on developing an integrated recommendation engine built upon two complementary approaches: similarity matrices and propensity models. While El Abbass's comprehensive framework encompasses five pillars (Recommendable Products, Similarity Matrix, Propensity Models, Website Visits, and Communication Profile), this implementation concentrates on the core recommendation generation components that align with the Bank's available data and strategic needs.

The recommendation engine developed in this study consists of two interconnected pillars:

1. **Similarity-Based Recommendation Engines** is a dual-perspective approach to pattern recognition in banking relationships with two types.

    - *Customer Similarity Engine*: Identifies customers with similar product ownership patterns and recommends products that similar customers own but the target customer does not, following the "customers like you also bought" principle.

    - *Product Similarity Engine*: Analyzes products frequently co-owned by customers and recommends complementary products based on existing customer ownership, following the "customers who bought this also bought" principle.

2. **Propensity Models** are models that determine the likelihood of a customer adopting specific products based on their demographic attributes and existing banking relationships, implemented using XGBoost for each product category.

These pillars are integrated through a weighted recommendation approach. Each recommendation source (customer similarity, product similarity, and propensity models) generates a ranked list of candidate products with corresponding scores. The final recommendation list is produced by computing a weighted sum of these scores for each candidate product:

$$\text{Final Recommendations} = \mathrm{X}\mathrm{y}_1\mathrm{y}_2\mathrm{y}_3$$

where:

- X is the weight matrix containing product-specific weights for each recommendation source
- y_1 is the vector of scores generated by the Customer Similarity model
- y_2 is the vector of scores generated by the Product Similarity model
- y_3 is the vector of scores generated by the Propensity model

This matrix formulation allows for efficient computation of the final weighted scores across all products and recommendation sources. After calculating these combined scores, the system selects the top-N products with the highest scores as the final recommendations for the customer, where N is a configurable parameter that determines the number of suggestions to present.

The sections that follow detail the implementation of each component, including data preprocessing, mathematical foundations, algorithmic approaches, and integration methods. The system is evaluated using a consistent train-test split methodology, ensuring that all components are assessed under comparable conditions.

3.1 Dataset

The dataset used in the paper consists of two primary components: a sparse product matrix and a customer attribute dataset. It contains data for 15546 customers and 30 different features which consists of different product flags and customer attributes. For data privacy reasons, the dataset the bank gave is a random subset of their customer base, so it may or may not reflect the actual distribution. By combining product ownership data with customer attributes, the dataset enables the NBO system to identify patterns in product adoption and customer behavior. This merged dataset serves as the foundation for both the propensity model, while the similarity matrix only relies on the sparse matrix to drive the recommendation process.

3.1.1 Product Sparse Matrix

This sparse matrix indicates product ownership across the Bank's offerings for a sample of actual customers. Each row represents a customer, while each column represents a specific product, with binary values indicating whether a customer owns that product which serves as the foundation for generating recommendations. Institutional Banking (IB), however, is an integer value which reflects the count of IB products owned by the client.

Description of Product Features

The products included in the matrix are grouped into the following categories:

1. `CORPORATE_LOANS`: Encompasses loans for business purposes, including trade loans, leasing, term loans, and other specialized financing options.

2. `RETAIL_LOAN`: Refers to retail loan products aimed at individual customers, such as group plans, fleet financing, and contract-to-sell financing.

3. `TRADE_SERVICES`: Includes trade-related financial services like letters of credit, standby letters of credit (SBLC), and duties payment services for businesses.

4. `INSURANCE`: Covers various insurance products, including life and non-life policies, that help protect customer assets and manage risks.

5. `CORPORATE_FINANCE`: Encompasses corporate-level financial services such as underwriting equity, debt, project finance, structured finance, and advisory services.

6. `FOREX`: Refers to foreign exchange services, indicating customer involvement in currency trading or conversions.

7. `CORPORATE_CARDS`: Includes corporate credit cards, both credit and prepaid, for business expenses and purchases.

8. `BB`: Stands for Business Banking, which includes financial products and services tailored to business clients, such as working capital loans and business deposits.

9. `DEPOSITS`: Refers to deposit accounts, commonly maintained by customers for managing cash and transactions, including business banking (BB) and other savings types.

10. RETAIL_PRODUCTS: Encompasses a variety of retail-oriented loan products, such as auto loans, housing loans, and personal loans, catering to individual customer needs.

11. INVESTMENTS_AND_SECURITIES: Represents a broad range of investment products, including both fixed income and equity securities, aimed at growing customer wealth.

12. REMITTANCE: Represents remittance services, commonly used by customers for sending or receiving funds internationally.

13. BUILDUP_F: Involves financial products designed to increase the value of customer holdings over time, such as savings plans and structured deposits.

14. CASH_MANAGEMENT: Refers to services that help businesses manage their cash flow, including payment processing, liquidity management, and treasury services.

15. INVESTMENT_FUNDS: Represents mutual funds or segregated portfolios for customers seeking to grow their wealth over time.

16. SECURITIES: Involves investments in fixed income and equity securities, often associated with higher-level investment activities.

17. IB: Count of IB products owned by the client.

Exploratory Data Analysis for Product Dataset

Figure 3.1 presents the class distribution across product categories. Distinction is shown between customers who own a product (Class 1) and those who do not (Class 0). Figure 3.1a illustrates the distribution for individual clients, Figure

3.1b for organizational clients, and Figure 3.1c combines both datasets to provide a comprehensive view of product ownership patterns across all customer segments.

We can observe that there is a class imbalance with certain products demonstrating higher adoption rates among specific client types. BB (Business Banking) and Deposits show the highest adoption rates for both type of clients which indicates that these products are widely held across customer segments. In contrast, products such as Corporate Finance, Corporate Cards, and Trade Services are severely imbalanced. These observations indicates that there are only few universally popular financial products, and most are specialized and caters to a specific client needs.

Table 3.1 shows the summary statistics for IB products, segmented by the entire dataset and across the two different client groups.

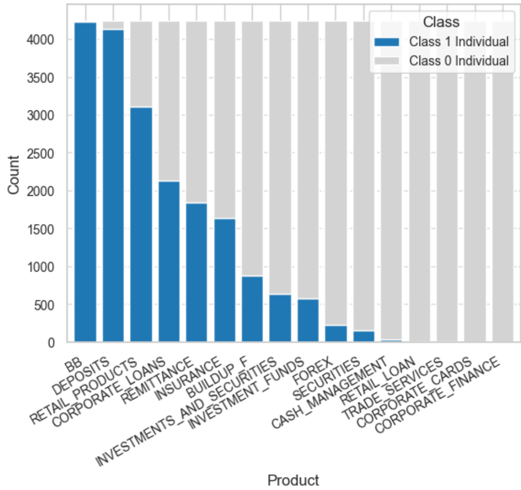
Statistic	**Individual**
Count	4,246
Mean	3.46
Std Dev	2.15
Min	0
25th Percentile	2
50th Percentile	3
75th Percentile	5
Max	16

Statistic	**Organization**
Count	11,300
Mean	1.66
Std Dev	1.21
Min	0
25th Percentile	1
50th Percentile	1
75th Percentile	2
Max	10

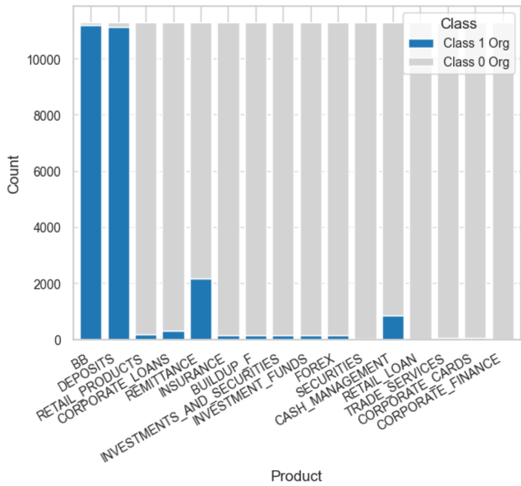
Statistic	**Combined**
Count	15,546
Mean	2.15
Std Dev	1.72
Min	0
25th Percentile	1
50th Percentile	1
75th Percentile	3
Max	16

Table 3.1: Summary statistics for Institutional Banking (IB) product ownership across individual, organizational, and combined clients.

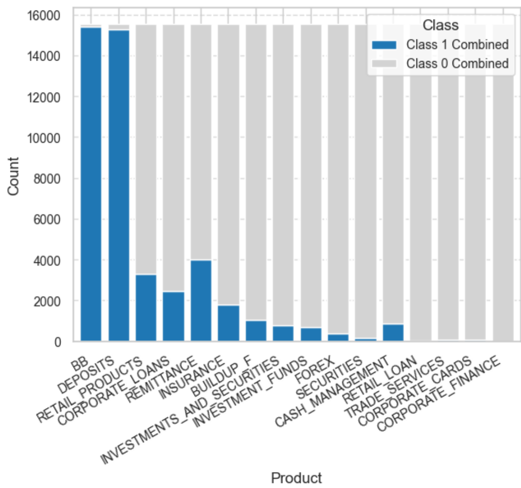
The summary statistics for IB product ownership reveal distinct patterns between individual and organizational clients. Individual clients have a higher average IB ownership of 3.46 with greater variance, suggesting broader prod-



(a) Individual Clients



(b) Organizational Clients



(c) Combined

Figure 3.1: Product Ownership Distribution across Individual, Organizational, and Combined Clients

uct engagement. In contrast, organizational clients exhibit lower average IB ownership of 1.66 with a smaller spread, indicating more consistent product adoption. Combined data shows an average of 2.15, with a wider range driven by individual client behavior. This suggests that individual clients are more diverse in their IB product adoption compared to organizational clients, who tend to engage with fewer products.

3.1.2 Customer Information Dataset

This dataset provides demographic and behavioral information on the Bank's customers. Key attributes include age, income source, region, business ownership, digital engagement flags, and other demographic data essential for the predictive modeling of product adoption.

Description of Client Features

The following are the features associated with the clients:

1. `CUST_NUM` (Float): A unique identifier assigned to each client, used to combine datasets and ensure data consistency.

2. `PROVINCE` (String): The province within the Philippines where the client's declared address is located.

3. `REGION` (String): The geographical region within the Philippines where the client's declared address is located.

4. `EDUCATION` (String): The highest level of educational attainment declared by the client.

5. `AGE` (Float): The client's age, measured in years.

6. `TENURE` (Float): The duration of the client's relationship with the bank, measured in years.

7. `BUSINESS_OWNER` (String): A binary indicator denoting whether the client is a business owner, based on self-declaration or the availing of business loans.

8. `DIGITAL_FLAG` (String): Indicates the client's preferred banking channel(s) either as Traditional or Digital.

9. `INCOME_SOURCE` (String): The source of income as declared by the client.

10. `IDV_OR_ORG` (String): The client type, classified as either an Individual or an Organization.

11. `SEGMENT` (String): The client's banking segment, determined based on the amount of money held in the bank over a specified period of time.

12. `GENDER` (String): The client's declared gender.

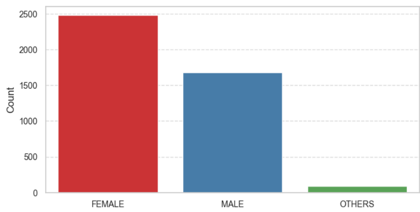
Exploratory Data Analysis for Client Features

This section presents an exploratory analysis of the client features to understand the underlying patterns and characteristics of the customer base. The analysis covers both categorical and numerical features, providing insights into customer demographics and behavioral trends. Since organizational clients do not have values for these features, the analysis is limited to the individual clients.

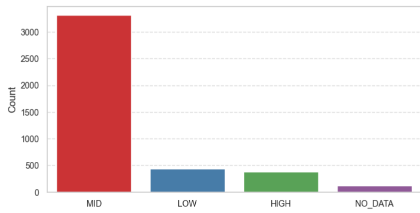
Categorical Feature Distributions

The categorical feature distributions provide insights into the demographic and behavioral characteristics of individual clients. This section examines key categorical attributes, including gender, education level, business owner status. income source, and region.

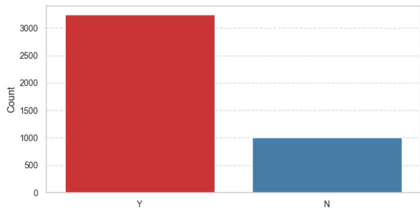
Figure 3.2 presents visualizations for three categorical features of individual clients. Figure 3.2a shows the gender distribution, Figure 3.2b for educational status, and Figure 3.2c for business owner status.



(a) Gender Distribution



(b) Education Level Distribution



(c) Business Owner Status

Figure 3.2: Gender, Education Level, and Business Owner Status Distributions for Individual Clients

Analysis of the dataset demonstrates a gender distribution favoring female clients, with a substantial proportion of customers possessing mid-level education. Furthermore, a majority of the client base is comprised of business owners.

Figure 3.3 presents the distribution of individual clients across various regions in the Philippines.

The majority of clients are concentrated in the National Capital Region (NCR), followed by Region IV-A (CALABARZON) and Region III (Central Lu-

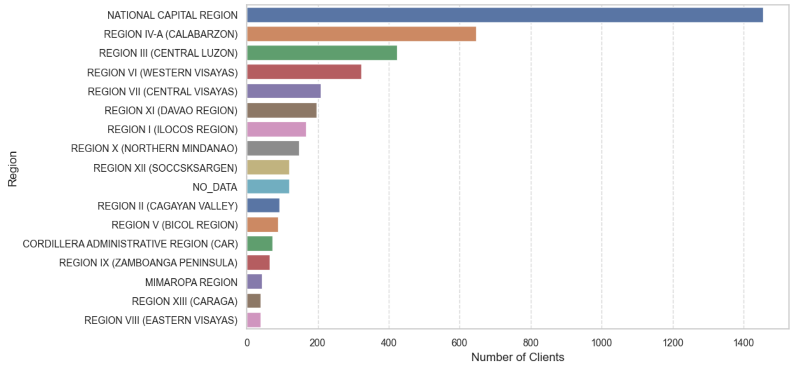


Figure 3.3: Region Distribution for Individual Clients

zon), highlighting stronger customer engagement in urbanized areas. On the other hand, regions like CARAGA, Eastern Visayas, and MIMAROPA have the lowest client counts, indicating lower market penetration in these areas. The presence of "NO_DATA" suggests missing regional information for a small portion of clients.

Figure 3.4 illustrates the distribution of income sources among individual clients.

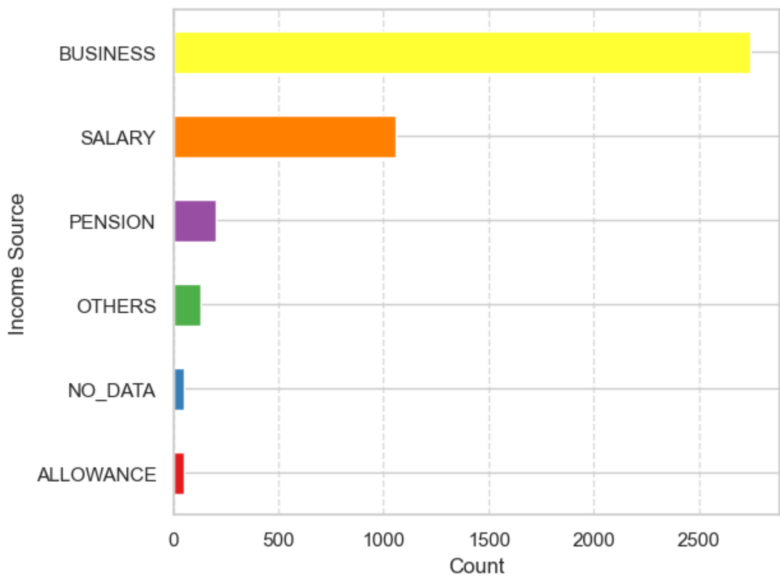


Figure 3.4: Income Source Distribution for Individual Clients

Most clients derive their income from business and salaries, representing the largest share of the customer base. Pensions and allowances contribute to a smaller but notable portion. The "NO_DATA" category reflects incomplete income information for some clients. A collection of less frequent income sources, including remittances, commissions, unspecified other sources, interest from savings, placements, and investments, as well as economically inactive individuals, were aggregated into the "Others" category due to their relatively low representation within the dataset.

Summary Statistics for Numerical Features

This section examines key numerical attributes of individual clients, including age and tenure, highlighting their central tendencies and variability to better understand customer profiles and product engagement patterns.

Table 3.2 presents the summary statistics for the numerical features among individual clients.

Statistic	AGE	TENURE
Count	4,220	4,241
Mean	52.01	15.58
Standard Deviation	13.19	7.64
Minimum	17.00	0.32
25th Percentile	42.00	9.17
Median (50th Percentile)	51.00	15.13
75th Percentile	62.00	23.00
Maximum	98.00	31.15

Table 3.2: Summary Statistics for Numerical Features (Individual)

From Table 3.2, we can observe that the average customer age is 52.01 years, with a median of 51 years suggests that the majority of clients are mid-to-late career individuals. The average tenure is 15.58 years, with a median of 15.13 years indicates a long-standing relationship with the bank. These statistics highlight the age and tenure distribution, which can inform customer segmentation and personalized strategies.

3.2 Data Preprocessing

This section contains the steps undertaken to clean and prepare the dataset for model training and evaluation. It includes the handling of missing values, normalization of variables, and the splitting of data into training and test sets.

3.2.1 Handling Missing Values

Banking datasets frequently contain missing values due to incomplete customer information or varied data collection across channels. In Table 3.2, 26 individual customer records exhibited missing AGE values, and 5 records lacked DNA_TENURE_NUM values. These missing values were imputed using the respective median values for each variable. For organizational customers, we removed individual-specific columns including PROVINCE, REGION, EDUCATION, DNA_AGE, DNA_TENURE_NUM, BUSINESS_OWNER, DIGITAL_FLAG, and INCOME_SOURCE as these attributes are not applicable to them.

3.2.2 Normalization of IB Feature

The Institutional Banking (IB) variable requires special handling as it is of the integer data type. Unlike the binary product indicators, this count variable needed normalization. For each customer segment (individual and organizational), the maximum IB value was determined, and all IB values were divided by this maximum, creating an IB_NORM feature ranging from 0 to 1.

3.2.3 Train-Test Split for Model Training and Evaluation

After cleaning the data, the dataset was split into training and testing sets using an 80-20 split. The dataset was first divided by customer type (individual and organizational) for both the propensity and similarity-based models. Within each customer type, 80% of the data was allocated to the training set and 20% to the test set. This ensured that the models were trained and evaluated on independent sets, providing a reliable foundation for performance assessment.

3.3 Pillar 1: Similarity Matrix

The Similarity Matrix is a core pillar of the Next Best Offer (NBO) system that identifies patterns in customer product ownership through mathematical techniques rather than relying on subjective rules or departmental silos.

This study implements a dual approach to similarity-based recommendations:

1. Customer Similarity System: Using the principle "customers like you also bought," this approach identifies similar customer profiles and recommends products they own that the target customer does not. This builds on collaborative filtering techniques established by Sarwar et al. [22], who demonstrated that customers with similar purchasing patterns often share underlying preferences.

2. Product Similarity System: Following El Abbass's item-based methodology [1], this approach operates on the principle "customers who bought this also bought." It identifies products frequently co-owned by customers and recommends complementary products based on the customer's existing portfolio.

This dual perspective creates a comprehensive framework that captures both customer-to-customer relationships and product-to-product affinities, providing the Bank with a robust foundation for personalized recommendations and addressing the limitations of its previously fragmented approach.

3.3.1 Choosing the Distance Metric

The effectiveness of the similarity matrix fundamentally depends on how accurately we can quantify relationships between entities. The choice of distance metric directly impacts which customers are considered similar and which products are recommended, making it one of the most critical decisions in the NBO system design.

Distance Metrics Considered

The following definitions introduce several common similarity and distance metrics used to quantify the relationship between vectors or sets, which can be used to construct a similarity matrix.

Definition 1 (Cosine Similarity). *Let A_i and B_i be the i-th components of vectors A and B, respectively, where n is the dimension of the vectors. The cosine similarity between two vectors, A and B, measures the cosine of the angle between them and is defined as:*

$$\mathit{cosine\_similarity}(A, B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Definition 2 (Jaccard Similarity). *The Jaccard similarity between two sets A and B is defined as the ratio of the cardinality of their intersection to the cardinality of their union:*

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Definition 3 (Hamming Similarity). *Let A and B be two binary vectors of equal length n, where A_i and B_i represent the i-th elements of the vectors. The Hamming distance between A and B measures the number of positions at which the corresponding elements differ and is defined as:*

$$\mathit{Hamming\_distance}(A, B) = \sum_{i=1}^{n} |A_i - B_i|$$

The Hamming similarity measures the proportion of matching elements between two vectors, with a similarity of 1 indicating identical vectors and a similarity of 0 indicating complete dissimilarity. It is derived from the Hamming distance through normalization where n is the total number of products.:

$$\mathit{Hamming\_similarity}(A, B) = 1 - \frac{\mathit{Hamming\_distance}(A, B)}{n}$$

Selection of Jaccard Similarity

While all three metrics were implemented and tested with the Bank's data, Jaccard similarity was selected as the primary distance metric for several key reasons. The Bank's product ownership data is primarily binary (customers either own a product or they don't) with the exception of the IB count, and Jaccard is specifically designed for comparing sets, making it naturally aligned with this data structure. The Jaccard score also provides an easily interpretable measure—the proportion of shared products relative to the total unique products—which aligns with how banking professionals naturally think about customer portfolios. In preliminary testing, Jaccard similarity consistently produced more intuitive customer groupings and more relevant recommendations than the alternatives when applied to the Bank's specific dataset. Additionally, for sparse binary data, Jaccard calculations can be optimized to focus only on non-zero elements, providing performance benefits for large-scale recommendation systems.

The NBO system implementation nevertheless maintains flexibility to use any of the three metrics, allowing for different approaches based on specific recommendation scenarios or future data characteristics. For institutional banking products with count values rather than binary indicators, the system can employ cosine similarity which better handles non-binary values.

3.3.2 Customer Similarity System

The Customer Similarity System implements a "customers like you also bought" approach to recommendations, leveraging patterns in product ownership across similar customer segments. This system identifies customers with similar product portfolios to a target customer and recommends products they own that the target customer does not.

Computing Customer Similarity

The system represents each customer as a vector consisting of binary product ownership indicators and an Institutional Banking (IB) count. Each feature, including both product ownership indicators and the IB count, is given equal weight in the similarity calculation.

When computing similarity using distance metrics like Jaccard, the system first normalizes the IB count to ensure comparability with the binary product features. This ensures that the similarity calculation is balanced across all features regardless of their original scale.

Computational Efficiency Considerations

Computing pairwise similarity between all customers in a large dataset would be computationally expensive. While the initial training process does calculate the full customer similarity matrix for analysis purposes, the recommendation process uses a more efficient approach. For a given target customer, the system computes similarity directly against all other customers using vectorized operations, applying a similarity threshold to retain only meaningful relationships. This optimization enables real-time recommendations even for large customer bases.

Generating Recommendations

Once similar customers are identified, the system calculates a weighted score for each candidate product not owned by the target customer:

$$\text{Score}_{customer} = \frac{\sum_{i=1}^{N} S(C, S_i) \cdot I_{S_i}(p)}{\sum_{i=1}^{N} I_{S_i}(p)}$$

where:

- $S(C, S_i)$ is the similarity score between the target customer C and similar customer S_i
- $I_{S_i}(p)$ is an indicator function that equals 1 if similar customer S_i owns product p, and 0 otherwise
- N is the total number of similar customers

This approach weights each similar customer's contribution by their similarity score, giving more influence to customers who more closely match the target customer's profile. Products are then ranked by their scores, with higher scores indicating stronger recommendations.

3.3.3 Customer Similarity Sample Calculation

To illustrate the customer similarity calculation and recommendation process, consider a simplified example with four customers (C1-C4) and three product categories (A, B, C), plus an Institutional Banking (IB) count for similarity calculation:

Customer	Product A	Product B	Product C	IB Count
C1 (Target)	1	0	0	1
C2	1	1	0	2
C3	0	0	1	0
C4	1	0	1	3

Table 3.3: Simplified customer-product ownership matrix

In Table 3.3, customer C1 is our target customer who currently owns Product A and has an IB count of 1.

Step 1: Calculate similarity between C1 and all other customers using Jaccard similarity

For binary product features (A, B, C):

$$J(C1, C2) = \frac{|C_1 \cap C_2|}{|C_1 \cup C_2|} = \frac{1}{2} = 0.5$$
$$J(C1, C3) = \frac{|C_1 \cap C_3|}{|C_1 \cup C_3|} = 0$$
$$J(C1, C4) = \frac{|C_1 \cap C_4|}{|C_1 \cup C_4|} = 0.5$$

For IB count, we first normalize the values to [0,1] by dividing by the maximum (3):

$$IB_{norm}(C1) = \frac{1}{3} = 0.33$$
$$IB_{norm}(C2) = \frac{2}{3} = 0.67$$
$$IB_{norm}(C3) = \frac{0}{3} = 0$$
$$IB_{norm}(C4) = \frac{3}{3} = 1$$

Then convert the normalized value to 1 if it is greater than 0, and compute the Jaccard similarity accordingly:

$$J_{IB}(C1, C2) = \text{For binary feature: } \frac{1}{1} = 1 \text{ (both > 0)}$$
$$J_{IB}(C1, C3) = \text{For binary feature: } \frac{0}{1} = 0$$
$$J_{IB}(C1, C4) = \text{For binary feature: } \frac{1}{1} = 1 \text{ (both > 0)}$$

Step 2: Compute overall similarity (equal weighting)

We now combine the Jaccard similarity for the binary products and IB count by assigning weights to each one of them that sums up to one. For this example, we use 0.5 for both binary and IB weights.

$$S(C1, C2) = 0.5(0.5) + (0.5)(1) = 0.75$$

$$S(C1, C3) = 0.5(0) + 0.5(0) = 0$$

$$S(C1, C4) = 0.5(0.5) + 0.5(1) = 0.75$$

Step 3: Identify similar customers (threshold = 0.1)

Customers C2 and C4 both exceed our similarity threshold (0.75 > 0.1) and are considered similar to C1. Customer C3's similarity (0) does not exceed the threshold, so it is excluded from consideration.

Step 4: Generate product recommendations

C1 currently owns Product A but not B or C. Let's calculate scores for these missing products:

For Product B:

$$\begin{aligned} \text{Score}_B &= \frac{S(C1, C2) \cdot I_{C2}(B) + S(C1, C3) \cdot I_{C3}(B) + S(C1, C4) \cdot I_{C4}(B)}{I_{C2}(B) + I_{C3}(B) + I_{C4}(B)} \\ &= \frac{0.75 \cdot 1 + 0.25 \cdot 0 + 0.75 \cdot 0}{1 + 0 + 0} = 0.75 \end{aligned}$$

For Product C:

$$
\begin{aligned}
\text{Score}_C &= \frac{S(C1, C2) \cdot I_{C2}(C) + S(C1, C3) \cdot I_{C3}(C) + S(C1, C4) \cdot I_{C4}(C)}{I_{C2}(C) + I_{C3}(C) + I_{C4}(C)} \\
&= \frac{0.75 \cdot 0 + 0.25 \cdot 1 + 0.75 \cdot 1}{0 + 1 + 1} = \frac{1}{2} = 0.5
\end{aligned}
$$

Final Recommendation for C1:

- Recommended Products: B (score 0.75), C (score 0.5)
- Similar customers found: 3 (C2, C3, and C4, with C2 and C4 being most similar)

This example demonstrates how the system identifies relevant products based on similar customers' portfolios, providing tailored recommendations based on observed product ownership patterns.

3.3.4 Product Similarity System

While the Customer Similarity System finds relationships between customers, the Product Similarity System identifies connections between products based on co-ownership patterns. This approach follows the principle "customers who bought this also bought," providing recommendations based on products that tend to be owned together across the customer base.

Computing Product Similarity

To compute product similarity, the system first transposes the customer-product ownership matrix. In this transposed matrix, each row represents a product rather than a customer, and each column represents a customer. The value at each position indicates whether a customer owns that product (1) or not (0).

This transposition transforms the perspective from customer-centric to product-centric:

- In the original matrix, each row is a customer vector showing which products they own
- In the transposed matrix, each row is a product vector showing which customers own it

Using this transposed matrix, the system calculates similarity between products using the same distance metrics discussed earlier (Jaccard, Cosine, or Hamming). For example, the Jaccard similarity between two products P_1 and P_2 is calculated as:

$$J(P_1, P_2) = \frac{|\text{Customers owning both } P_1 \text{ and } P_2|}{|\text{Customers owning either } P_1 \text{ or } P_2|}$$

The resulting product similarity matrix captures how frequently products are co-owned across the customer base. Higher similarity scores indicate products that often appear together in customer portfolios, suggesting potential complementarity or related customer needs.

Threshold Filtering

To reduce noise and focus on meaningful relationships, the system applies a similarity threshold. Product pairs with similarity scores below this threshold are treated as having zero similarity. This filtering step:

- Eliminates weak or potentially coincidental relationships
- Reduces computational complexity for recommendation generation
- Improves recommendation quality by focusing on stronger patterns

The threshold is configurable based on the specific characteristics of the Bank's product portfolio and customer base.

Generating Recommendations

The product-based recommendation process first identifies which products the target customer already owns. Unlike the Customer Similarity System, which works even for new customers without products, the Product Similarity System requires at least one owned product to generate recommendations.

For each candidate product not owned by the customer, the system calculates an average similarity score based on the customer's existing product portfolio:

$$\text{Score}_{product} = \frac{\sum_{i=1}^{n} S(P, O_i)}{n}$$

where:

- $Score_{product}$ is the overall similarity score for candidate product P
- $S(P, O_i)$ is the similarity score between the candidate product P and an owned product O_i
- n is the number of products the customer already owns

This calculation provides an average similarity between the candidate product and the customer's existing portfolio. Products with higher average similarity scores are considered more relevant for recommendation.

Final Recommendation Output

The Product Similarity System generates recommendations in a structured format that includes:

- A ranked list of products sorted by similarity score
- The similarity score for each recommended product

- The customer's current product portfolio used as the basis for recommendations

If a customer owns no products or if no products exceed the similarity threshold, the system returns an appropriate message indicating that no product-based recommendations can be generated.

3.3.5 Product Similarity Sample Calculation

To illustrate the Product Similarity approach, we'll use the same simplified data from our previous example, but now focusing on product-to-product relationships as illustrated in Table 3.4 before transposing them to get the product vectors as seen in Table 3.5

Customer	Product A	Product B	Product C
C1	1	0	0
C2	1	1	0
C3	0	0	1
C4	1	0	1

Table 3.4: Customer-product ownership matrix

Step 1: Transpose the matrix to get product vectors

Product	C1	C2	C3	C4
A	1	1	0	1
B	0	1	0	0
C	0	0	1	1

Table 3.5: Transposed product-customer matrix

Step 2: Calculate Jaccard similarity between products

$$J(A,B) = \frac{|\{C2\}|}{|\{C1,C2,C4\}|} = \frac{1}{3} \approx 0.33$$
$$J(A,C) = \frac{|\{C4\}|}{|\{C1,C2,C3,C4\}|} = \frac{1}{4} = 0.25$$
$$J(B,C) = \frac{|\{\}|}{|\{C2,C3,C4\}|} = \frac{0}{3} = 0$$

Step 3: Construct the product similarity matrix

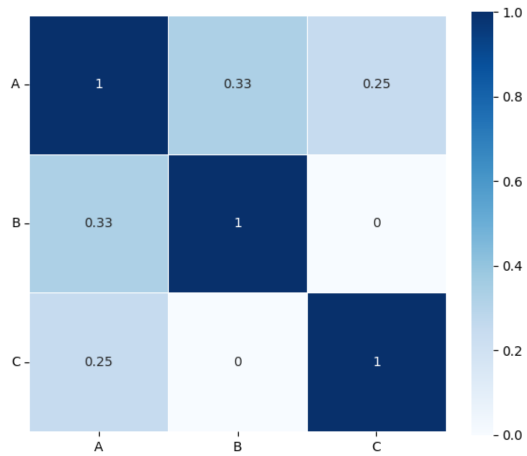


Figure 3.5: Product similarity matrix

Step 4: Generate recommendations for a customer

Consider customer C1 who owns only Product A. We calculate similarity scores for products they don't own:

$$\text{Score}_B = \frac{S(B,A)}{1} = \frac{0.33}{1} = 0.33$$
$$\text{Score}_C = \frac{S(C,A)}{1} = \frac{0.25}{1} = 0.25$$

Final Product-Based Recommendation for C1:

- Recommended Products: B (score 0.33), C (score 0.25)

- Based on: Ownership of Product A

This example demonstrates how the Product Similarity System identifies relationships between products based on co-ownership patterns, as seen in Figure 3.5, and leverages these relationships to generate recommendations for customers based on their existing portfolio.

3.4 Pillar 2: Propensity Models

While the similarity matrix approaches recommendations from patterns of co-ownership and customer likeness, the Propensity Models directly predict the likelihood of a customer adopting specific products based on their demographic attributes and existing product relationships. This prediction-based approach complements the pattern-matching nature of similarity matrices by capturing complex, non-linear relationships between customer characteristics and product adoption.

These product-specific propensity scores are then used to generate recommendations by ranking products according to their predicted adoption likelihood. Products that a customer already owns are excluded, and the remaining products with the highest propensity scores form the final recommendations. This approach directs marketing efforts toward products with the highest statistical likelihood of customer adoption.

The implementation uses XGBoost, a gradient boosting framework known for its efficiency and effectiveness with structured data. The system handles individual and organizational customers separately, recognizing the fundamental differences in their banking needs and adoption patterns.

3.4.1 Mathematical Foundations of XGBoost

This subsection is based on the work of Chen and Guestrin [7], which provides the foundational mathematics behind XGBoost. XGBoost, or Extreme Gradient Boosting, is a tree-based ensemble method that builds models sequentially in an additive manner, combining the strengths of multiple decision trees to minimize prediction error. The algorithm's foundation lies in gradient boosting, where each new tree attempts to correct the errors made by previous trees. The algorithm's foundation lies in gradient boosting, where each new tree attempts to correct the errors made by the previous trees [7].

Objective Function

The objective function in XGBoost consists of two main components: a loss function that measures the model's predictive accuracy and a regularization term that penalizes model complexity. The general form of the objective function is:

$$\text{Obj} = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where:

- $L(y_i, \hat{y}_i)$ is the loss function that computes the error between the predicted value $\hat{y}_i$ and the true label y_i.
- $\hat{y}_i$ represents the predicted value for the i-th data point.
- y_i represents the true label for the i-th data point.
- n is the total number of data points.
- $\Omega(f_k)$ is the regularization term for the k-th tree f_k.
- f_k represents the k-th tree in the ensemble.

- K is the total number of trees in the ensemble.

For binary classification, XGBoost commonly uses the logistic loss function, defined as:

$$L(y, \hat{y}) = -\left(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\right)$$

where:

- y is the true label.
- $\hat{y}$ is the predicted probability of the positive class.

Additive Learning and Gradient Descent

XGBoost builds trees in a sequential manner by adding a new tree at each iteration to improve the model as seen in Figure 3.6. Specifically, the algorithm minimizes the residual error from the previous trees by fitting a new tree that predicts these residuals. This process is mathematically equivalent to gradient descent, where the gradients of the loss function are used to adjust predictions incrementally.

At each iteration t, XGBoost defines the prediction for each instance i as:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i)$$

where:

- $\hat{y}_i^{(t)}$ is the predicted value for the i-th data point at iteration t.
- $\hat{y}_i^{(t-1)}$ is the predicted value for the i-th data point at iteration $t - 1$ (i.e., the prediction from the previous iteration).
- η is the learning rate (or shrinkage parameter).

- $f_t(x_i)$ is the output of the t-th tree for the i-th data point, evaluated at the feature vector x_i.

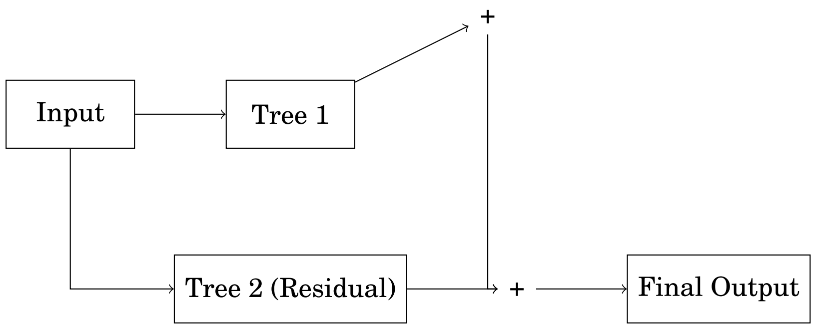


Figure 3.6: Illustration of Additive Learning in XGBoost

Tree Structure and Regularization

Each tree in XGBoost is constructed using a split-finding algorithm that minimizes the objective function, taking into account both the accuracy of predictions and the complexity of the tree. To prevent overfitting, XGBoost employs two forms of regularization:

- **L1 Regularization** on the weights of leaves, which promotes sparsity by assigning zero weights to unimportant nodes.
- **L2 Regularization** on the weights, penalizing large values to keep the model stable.

The combined effect of L1 and L2 regularization results in a more robust and generalized model, particularly useful for handling high-dimensional data and avoiding over-complex trees.

3.4.2 Additional Data Preparation

Feature Engineering

Feature engineering was applied differently for individual and organizational customers to reflect their distinct banking behaviors and needs. Moreover, feature engineering was performed separately for each target product model to prevent data leakage, ensuring that the target variable being predicted never influenced the features used for its prediction.

For both individual and organizational customers, several common features were created to capture product ownership patterns. A Total Products count was computed for each customer by adding the total product count and IB count, excluding the target product being predicted, to represent the breadth of their banking relationship. Additionally, an Investment Product Indicator was created as a binary flag indicating whether the customer owned any investment products (INVESTMENTS_AND_SECURI-TIES, INVESTMENT_FUNDS, or SECURITIES) to help in identifying clients with investment potential.

Individual customer features emphasized demographic and behavioral dimensions to capture life-stage and financial sophistication. Age-based features included binary indicators such as `is_senior` as created to indicate whether a customer is over 60 years old, capturing potential age-related behavioral patterns. To capture relationship maturity effects, customer tenure was categorized into meaningful groups: new (0–1 years), early (1–5 years), established (5–10 years), loyal (10–20 years), and veteran (20+ years). Financial status was represented through binary flags for business ownership and digital channel usage. Income source and education level were transformed into ordinal features through hierarchical mappings, grouping these categorical variables by economic significance. The individual customer feature set also included product affinity gaps identifying potential adoption opportunities based on customer characteristics. These included flags for high-income customers without investment products, business owners without loan products, digital channel users without remittance services, and senior customers without insurance products.

For organizational customers, complexity and relationship features were emphasized to capture business banking needs. Business product groupings were created as binary indicators for categories such as loan products, transaction products, and corporate products to reflect the functional roles these services in the organization. A Product Complexity Score was calculated as the ratio representing the proportion of available products owned by each organization, serving as a proxy for relationship complexity. The organizational customer feature set included gap indicators for businesses with deposits but without cash management services, those using remittance services but not foreign exchange, organizations with corporate loans but without corporate finance services, and complex banking relationships without trade services. These features are relevant as they enable more targeted product recommendations for the organizations based on their needs.

Feature Encoding and Model Preparation

After feature engineering, the data required additional preprocessing to prepare it for the XGBoost algorithm. This involved appropriate encoding of categorical features and addressing class imbalance issues. Categorical features were handled based on whether they possessed inherent ordering. For variables with meaningful ordinal relationships, numerical mapping was applied. EDUCATION was converted to a three-point scale where 'LOW' mapped to 1, 'MID' to 2, and 'HIGH' to 3. Similarly, INCOME_SOURCE was transformed based on economic significance, with categories like 'PENSION' and 'ALLOWANCE' receiving lower values (1), salaried income at a middle value (2), and business or investment income assigned higher values (3). Customer tenure (DNA_TENURE_NUM) was grouped into sequential categories representing relationship maturity: new, early, established, loyal, and veteran.

For categorical variables such as PROVINCE, REGION, SEGMENT, and GENDER, one-hot encoding was applied to convert them into binary indicator columns. This transformation expanded the feature space but preserved the non-ordinal nature of these variables. Banking datasets typically show significant class imbalance, with many products having low adoption rates.

For product categories with severe imbalance (CORPORATE_FINANCE, CORPORATE_CARDS, and TRADE_SERVICES), Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training data. SMOTE generated synthetic examples of the minority class based on the characteristics of existing adopters. This technique was particularly valuable for products with extremely low adoption rates, as it improved the model's ability to identify potential adopters without substantially increasing false positives. By applying SMOTE, the training set became more balanced, which enabled the models to better learn the decision boundary for rare product adoption.

3.4.3 Generating Recommendations

After training the individual product propensity models, the system employs them to generate personalized recommendations for customers. The recommendation process follows a systematic approach:

For each customer requiring recommendations, the system first identifies which products they already own. These existing products are excluded from recommendation consideration, as the goal is to suggest products the customer does not yet have.

Next, for each potential product not currently owned by the customer, the system:

1. Determines the appropriate propensity model based on the product category and customer type (individual or organizational)
2. Extracts the customer's demographic and product ownership features
3. Applies the same preprocessing and feature engineering steps used during training
4. Feeds the processed features into the corresponding propensity model
5. Obtains a propensity score representing the predicted probability (between 0 and 1) that the customer will adopt this product

Once propensity scores are calculated for all candidate products, they are ranked in descending order. Products with higher propensity scores are interpreted as having a greater likelihood of customer adoption and are therefore prioritized in the recommendation list.

The system then selects the top-N products with the highest propensity scores as the final recommendations, where N is a configurable parameter (typically set to 3-5 products).

3.5 Validation

The validation framework is essential for evaluating the effectiveness of our recommendation models and ensuring their real-world applicability. In the previous sections, we described three distinct recommendation approaches: customer similarity, product similarity, and propensity models. While each approach has its theoretical strengths, empirical validation is necessary to assess their practical performance and comparative effectiveness.

3.5.1 Leave-One-Out Validation Methodology

To rigorously evaluate the recommendation models, we implemented a leave-one-out validation strategy that simulates real-world recommendation scenarios. The approach consists of two key components:

First, we partitioned the dataset into training (80%) and test (20%) sets. The training set was used to construct the similarity matrices and train the propensity models, while the test set was reserved for unbiased performance evaluation. This split ensures that our models are evaluated on previously unseen data, providing a reliable measure of generalization capability.

Second, for each customer in the test set with multiple products, we systematically withheld one owned product at a time and treated it as the target for prediction. The recommendation system generated rankings based only on the remaining customer information and product ownership data. This simulates the practical scenario where the bank must recommend products to existing customers based on their incomplete portfolio, effectively testing whether our models can "rediscover" products that customers have already demonstrated interest in purchasing.

This methodology allows us to comprehensively assess not just whether a model can make recommendations, but whether those recommendations align with actual customer preferences as demonstrated by their existing product adoption decisions.

3.5.2 Evaluation Metrics

To provide a multifaceted evaluation of recommendation quality, we employed several complementary metrics:

Hit Rate@K

The Hit Rate@K measures the percentage of test cases where the withheld product appears within the top K recommendations. This metric directly addresses the question: "How often does our system recommend the product that the customer actually purchased?" We calculated this metric at K=1, K=3, and K=5, representing increasingly lenient evaluation criteria:

$$\text{Hit Rate@K} = \frac{\text{Number of times target product appears in top K recommendations}}{\text{Total number of test cases}}$$

A higher Hit Rate indicates better recommendation accuracy, with Hit Rate@1 being the most stringent measure of precision.

Mean Reciprocal Rank (MRR)

The Mean Reciprocal Rank provides insight into the average position of the withheld product within recommendation lists:

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

where $|Q|$ is the number of test cases and rank_i is the position of the withheld product in the recommendation list for the i-th test case. If the product does not appear in the recommendations, its reciprocal rank is 0.

MRR offers a more nuanced view than Hit Rate by rewarding systems that place the target product higher in the recommendation list. A value closer to 1 indicates that the target products consistently appear at or near the top of recommendations.

Normalized Discounted Cumulative Gain (NDCG)

NDCG evaluates the quality of the ranking by considering both the position and relevance of recommendations:

$$\text{NDCG@K} = \frac{\text{DCG@K}}{\text{IDCG@K}}$$

where DCG@K is the Discounted Cumulative Gain at position K:

$$\text{DCG@K} = \sum_{i=1}^{K} \frac{2^{\text{rel}_i} - 1}{\log_2(i+1)}$$

and IDCG@K is the DCG@K for the ideal ranking. In our leave-one-out validation, we assigned a relevance score of 1 to the withheld product and 0 to all others.

NDCG is particularly valuable for evaluating the overall quality of the ranking, as it penalizes relevant items appearing lower in the list more severely than hit rate metrics.

Precision@K

Precision@K measures the proportion of relevant recommendations among the top K:

$$\text{Precision@K} = \frac{\text{Number of relevant products in top K recommendations}}{K}$$

In our leave-one-out validation context, with only one relevant product per test case, Precision@K simplifies to $\frac{1}{K}$ when the target product appears in the top K, and 0 otherwise.

These complementary metrics provide a comprehensive assessment of recommendation quality. Hit Rate focuses on whether the target product appears at all, MRR emphasizes the rank position, NDCG considers the quality of the entire ranking, and Precision measures recommendation efficiency. Together, they allow us to evaluate different aspects of recommendation performance, from accuracy to ranking quality, providing a robust framework for comparing different models and optimizing their weights.

3.6 Ensemble Recommendation System

While each individual recommendation approach (customer similarity, product similarity, and propensity modeling) provides valuable insights, they each capture different aspects of the recommendation problem. The customer similarity system identifies patterns based on similar customer portfolios, the product similarity system recognizes co-ownership relationships, and propensity models predict adoption likelihood based on customer characteristics. An ensemble system that combines these complementary approaches can potentially deliver more robust and accurate recommendations than any single model alone.

To integrate the recommendations from multiple models, we developed a matrix-based approach that combines scores from each recommendation source using a configurable weight matrix. This approach allows different products to have different optimal weightings across the recommendation models.

For a given customer, the ensemble system first collects scaled recommendation scores from each component model:

- y_1: Vector of scaled scores from the customer similarity model
- y_2: Vector of scaled scores from the product similarity model
- y_3: Vector of scaled scores from the propensity model

These scores are then combined using a weight matrix $\mathbf{X}$, where each row corresponds to a product and each column corresponds to a model:

$$\text{Final Scores} = \mathbf{X} \cdot [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3]$$

Each element x_{ij} in the weight matrix represents the importance of model j for product i. The output of this calculation is a comprehensive list of products, each with a corresponding ensemble score that represents its predicted relevance to the customer. These final scores incorporate the weighted predictions from all three recommendation approaches, allowing the system to leverage the complementary strengths of each model in determining the most suitable products for recommendation.

3.6.1 Score Normalization

Before combining scores from different recommendation models, we needed to address the fact that each model produces scores on different scales with different distributions. We implemented min-max normalization to bring all scores into a consistent [0,1] range:

$$\text{normalized\_score} = \frac{\text{score} - \text{min\_score}}{\text{max\_score} - \text{min\_score}}$$

where min_score and max_score are calculated across all recommendations generated by a particular model. This normalization ensures that the weights applied to each model's scores have a consistent and meaningful impact regardless of the model's native scoring scale.

3.6.2 Weight Optimization

Determining the optimal weights for each product-model combination is a critical aspect of the ensemble system. The goal is to find weights that maximize our evaluation metrics, particularly hit rate, as this directly translates to capturing products that customers actually want. In a banking context, failing to recommend a product that a customer would adopt represents significant missed revenue opportunity, making hit rate optimization a business priority.

However, finding optimal weights presents a substantial computational challenge. With 13 product categories, 3 recommendation models, and weight increments of 0.1, the search space becomes overwhelming. For a comprehensive grid search of all possible product-specific weight combinations (where weights for each product sum to 1.0), we would need to evaluate approximately 66^{13} combinations (about 10^{23}) - a computationally infeasible task. Each weight combination would require recalculating recommendations and metrics for thousands of test cases.

To navigate this challenge, we explored several increasingly sophisticated optimization approaches:

Per-Model Weight Optimization

Our most straightforward approach assigned a single set of weights across all products, searching for values that maximized overall performance:

1. We defined a grid of weight combinations where weights sum to 1.0
2. For each combination, we applied the same weights to all products
3. We calculated aggregate performance metrics (hit rate, mean reciprocal rank, NDCG)
4. We selected the weight combination that maximized hit rate as recommendations are done as a list of products.

While computationally manageable (requiring only 66 evaluations), this approach failed to account for product-specific variations in model performance. Some models might excel at recommending certain product categories but perform poorly on others, a nuance lost with per-model weights.

Greedy Weight Matrix Optimization

Building on the per-model approach, we implemented a more sophisticated algorithm that considers interactions between products while remaining computationally feasible:

1. We started with an initial weight matrix (either equal weights or the per-product optimized weights)
2. We employed a greedy search with multiple random restarts to explore the weight space
3. For each restart:
   (a) Iteratively adjust weights for one product at a time
   (b) After each adjustment, we evaluated overall system performance
   (c) Keep changes that improved aggregate hit rate
   (d) Continue until no further improvements were found or reaching maximum iterations
4. Retain the best-performing weight matrix across all restarts

This approach sought to balance computational feasibility with optimization quality. The random restarts helped escape local maxima, while the greedy product-by-product optimization allowed us to efficiently search the high-dimensional weight space. To manage computational constraints, we used a step size of a 0.1 for weight adjustments and limited each search to 10 iterations with 3 random restarts.

Each of these approaches represents a different trade-off between optimization quality and computational complexity. In our final implementation, we employed the greedy weight matrix optimization approach, as it provided the best balance between computational feasibility and recommendation performance.

3.6.3 Final Recommendation Generation

After applying the optimized weights to normalized scores, the ensemble system:

1. Ranks all candidate products by their weighted scores
2. Excludes products the customer already owns
3. Selects the top-N products with the highest scores
4. Returns these as the final recommendations

This ensemble approach leverages the complementary strengths of different recommendation models, with the weight matrix dynamically adjusting the influence of each model based on the specific product being considered. The resulting recommendations benefit from both the pattern-recognition capabilities of similarity-based approaches and the predictive power of machine learning models.

CHAPTER IV

Results and Analysis

This chapter presents the comprehensive performance evaluation of the next best offer (NBO) system. We analyze the effectiveness of each recommendation component individually, followed by an assessment of the integrated ensemble system. Performance metrics are visualized to provide clear insights into the strengths and limitations of each approach.

4.1 Evaluation Framework

The NBO system was evaluated using a leave-one-out validation strategy on a test set of banking customers segmented into individual and organizational categories. The test set for individual customers consisted of 850 rows, while the organizational test set contained 2260 rows. For each customer, we systematically withheld one owned product and challenged the recommendation models to predict this withheld product. This approach simulates real-world usage where the system must recommend products that customers would genuinely be interested in acquiring.

4.1.1 Key Performance Metrics

We evaluated recommendation quality using several complementary metrics:

4.1.2 Key Performance Metrics

We evaluated recommendation quality using several complementary metrics:

- **Hit Rate**: The percentage of test cases where the withheld product category appears in the top 5 recommendations. This metric measures the system's overall ability to identify relevant products in a list of length 5.

  Example: If the INSURANCE category appeared in the recommended list in 8 out of 10 evaluations, regardless of its specific position, the Hit Rate@5 is 0.8.

- **Hit Rate@K**: The percentage of test cases where the withheld product category appears in the top K recommendations. We focus on K = 1 and K = 3 for position-sensitive evaluation. (Note: Hit Rate@5 is simply the overall Hit Rate defined above.)

  Example:

  If the INSURANCE product category appeared as the third product category in the recommendation list in 1 out of 10 evaluations, then this corresponds to a Hit Rate at 3 of 0.1.

- **Mean Reciprocal Rank (MRR)**: The average of the reciprocal ranks of the withheld product categories in the recommendation lists. This metric accounts for the position of the correct recommendation, with higher values indicating better performance.

  Example:

  Suppose the withheld product category INSURANCE appeared in the following positions across 10 test iterations:

$$\{2, 1, 3, 0, 4, 0, 1, 2, 5, 0\}$$

The corresponding reciprocal ranks are:

$$\left\{\frac{1}{2}, 1, \frac{1}{3}, 0, \frac{1}{4}, 0, 1, \frac{1}{2}, \frac{1}{5}, 0\right\}$$

Calculating the MRR:

$$\text{MRR} = \frac{1}{10}\left(\frac{1}{2} + 1 + \frac{1}{3} + 0 + \frac{1}{4} + 0 + 1 + \frac{1}{2} + \frac{1}{5} + 0\right) = \frac{1}{10}(3.78) = 0.378$$

This MRR of 0.3783 suggests that while the model often includes the correct product, it is ranked around the third as its MRR is around $\frac{1}{3}$.

- **Normalized Discounted Cumulative Gain (NDCG)**: Evaluates the quality of the ranking by considering both the position and relevance of recommendations, with diminishing returns for items appearing lower in the list.

  Example:

  Suppose the withheld product category INSURANCE appeared at rank 3 in the recommendation list of size 5. Since we only have one relevant item per test case (the withheld product), its relevance is 1, and the rest are 0.

$$\text{NDCG} = \frac{1}{\log_2(3+1)} = \frac{1}{\log_2(4)} = \frac{1}{2} = 0.5$$

  The ideal NDCG would place the relevant product at rank 1:

$$\text{IDCG@5} = \frac{1}{\log_2(1+1)} = \frac{1}{\log_2(2)} = 1$$

  This NDCG value indicates that the relevant item was found, but not in an ideal position.

- **Precision**: The proportion of relevant recommendations among all recommendations made, measuring the system's ability to provide accurate suggestions.

Example:

Suppose the INSURANCE product category is the correct recommendation. If it appears at any position within the top 5 recommendations in 6 out of 10 test cases, then:

$$\text{Precision} = \frac{6}{10 \times 3} = 0.2$$

This indicates that, on average, 20% of the top 3 recommendations are relevant.

These metrics collectively provide a comprehensive picture of recommendation quality, addressing both the presence of correct recommendations and their position in the recommended list.

4.2 Individual Models Performance

Before assessing the ensemble approach, we first analyze the performance of each individual recommendation model. This establishes a baseline for evaluating the ensemble system's effectiveness and identifies which individual model delivers the strongest recommendations.

4.2.1 Performance for Individual Customers

Customer Similarity Model Performance

The Customer Similarity model performance listed in Table 4.1 demonstrates moderate overall performance for individual customers, with a 90.76% hit rate and an overall HR@1 of 38.16%. This indicates that while the model can eventually identify relevant products (represented by the hit rate), it is less effective at ranking them at the top position.

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
INSURANCE	0.993865	0.516207	0.634382	0.198773	0.269939	0.720859	326
BB	0.839623	0.532115	0.609251	0.167925	0.336085	0.705189	848
DEPOSITS	0.875300	0.643465	0.701772	0.175060	0.485010	0.788969	834
RETAIL_PRODUCTS	0.942029	0.520719	0.625181	0.188406	0.294686	0.724499	621
INVESTMENTS_AND_SECURITIES	0.925926	0.869383	0.883904	0.185185	0.822222	0.918519	135
REMITTANCE	0.971831	0.573756	0.672614	0.194366	0.338028	0.684507	355
INVESTMENT_FUNDS	0.983607	0.954918	0.962430	0.196721	0.926230	0.983607	122
CORPORATE_LOANS	0.917453	0.580267	0.666436	0.183491	0.308962	0.801887	424
BUILDUP_F	0.891429	0.497238	0.597128	0.178286	0.217143	0.782857	175
FOREX	0.862745	0.343791	0.471705	0.172549	0.078431	0.607843	51
SECURITIES	1.000000	0.773504	0.831801	0.200000	0.588744	1.000000	39
CASH_MANAGEMENT	0.666667	0.222222	0.332476	0.133333	0.000000	0.444444	9
Overall	0.907591	0.583862	0.6649012	0.174392	0.381569	0.757044	3939

Table 4.1: Performance metrics for the Customer Similarity approach for Individual Customers.

The model shows significant variability across product categories. For Investment Funds, it achieves exceptional performance with both high hit rate (98.36%) and ranking accuracy (HR@1 of 92.62%, HR@3 of 98.36%). Similarly, for Investments and Securities, it achieves strong results with an HR@1 of 82.22%. These results suggest that customer similarity is particularly effective for investment products, likely because customers with similar financial profiles tend to adopt similar investment vehicles.

In contrast, the model struggles significantly with Cash Management (0% HR@1) and Forex (7.84% HR@1), despite achieving moderate hit rates for these products. This indicates that while the model can identify these products as relevant, it frequently fails to rank them in the top position.

The model's performance shows a relationship with product prevalence, but not as strongly as might be expected. While it performs relatively well for common products like BB and Deposits, its strongest performance is actually for some mid-frequency products like Investment Funds (122 test cases). This suggests that customer similarity captures specialized adoption patterns that transcend simple popularity effects.

Product Similarity Model Performance

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
INSURANCE	1.000000	0.670603	0.755168	0.200000	0.404908	0.957055	326
BB	1.000000	0.997642	0.998259	0.200000	0.995283	1.000000	848
DEPOSITS	1.000000	0.999400	0.999557	0.200000	0.998801	1.000000	834
RETAIL_PRODUCTS	1.000000	0.989533	0.992274	0.200000	0.979066	1.000000	621
INVESTMENTS_AND_SECURITIES	1.000000	0.500864	0.627019	0.200000	0.170370	0.866667	135
REMITTANCE	1.000000	0.753991	0.817564	0.200000	0.543662	1.000000	355
INVESTMENT_FUNDS	1.000000	0.474727	0.607811	0.200000	0.122951	0.909836	122
CORPORATE_LOANS	1.000000	0.853970	0.892042	0.200000	0.714623	0.997642	424
BUILDUP_F	1.000000	0.490952	0.618412	0.200000	0.165000	0.874286	175
FOREX	0.627451	0.200327	0.304165	0.125490	0.019608	0.274510	51
SECURITIES	0.846154	0.342735	0.466232	0.169231	0.102564	0.487179	39
CASH_MANAGEMENT	0.444444	0.088889	0.171935	0.088889	0.000000	0.000000	9
Overall	0.992384	0.857629	0.891963	0.198477	0.757553	0.966489	3939

Table 4.2: Performance metrics for the Product Similarity approach for Individual Customers.

The Product Similarity model (Table 4.2) demonstrates strong overall performance for individual customers, with a 99.24% hit rate, 75.76% HR@1, and 96.65% HR@3. This indicates that the model not only identifies relevant products but also ranks them highly.

The model excels with high-frequency products, achieving near-perfect results for BB, Deposits, and Retail Products, all with 100% hit rates and HR@1 values exceeding 97.9%. For these common products, co-ownership patterns are well-established, providing strong signals for recommendation.

However, performance drops significantly for specialized or less common products. Cash Management shows the poorest performance (44.44% hit rate, 0% HR@1 and HR@3), followed by Forex (62.75% hit rate, 1.96% HR@1). The low test counts for these products (9 and 51 respectively) suggest that limited co-ownership data affects the model's ability to detect meaningful patterns.

A notable pattern emerges with investment products: while the model achieves perfect hit rates for both Investments and Securities and Investment Funds, its ranking accuracy for these products is surprisingly low (HR@1 of 17.04% and 12.30% respectively). This indicates that the model recognizes these products as relevant but struggles to prioritize them appropriately, suggesting that investment product adoption may follow patterns less predictable by simple co-ownership analysis.

Propensity Model Performance

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
INSURANCE	1.000000	0.810583	0.859288	0.200000	0.656442	0.984663	326
BB	1.000000	1.000000	1.000000	0.200000	1.000000	1.000000	848
DEPOSITS	1.000000	0.998801	0.999115	0.200000	0.997602	1.000000	834
RETAIL_PRODUCTS	1.000000	0.959742	0.970246	0.200000	0.921095	1.000000	621
INVESTMENTS_AND_SECURITIES	1.000000	0.996296	0.997266	0.200000	0.992593	1.000000	135
REMITTANCE	1.000000	0.868873	0.902768	0.200000	0.754930	0.994366	355
INVESTMENT_FUNDS	0.983607	0.934426	0.947110	0.196721	0.893443	0.983607	122
CORPORATE_LOANS	1.000000	0.875904	0.907740	0.200000	0.773585	0.974057	424
BUILDUP_F	0.994286	0.667810	0.750963	0.198857	0.428571	0.931429	175
FOREX	0.960784	0.437582	0.565399	0.192157	0.196078	0.588235	51
SECURITIES	1.000000	0.544444	0.660350	0.200000	0.205128	0.897436	33
CASH_MANAGEMENT	1.000000	0.344444	0.502435	0.200000	0.111111	0.444444	9
Overall	0.998731	0.922340	0.941859	0.184392	0.862909	0.984259	3939

Table 4.3: Performance metrics for the Propensity Model approach for Individual Customers.

The Propensity Model (Table 4.3) demonstrates exceptional performance for individual customers, achieving a near-perfect overall hit rate of 99.87%, with strong ranking accuracy (HR@1 of 86.29% and HR@3 of 98.43%). This indicates that the model not only identifies relevant products but consistently ranks them at the top of recommendation lists.

For high-frequency products, the model achieves perfect or near-perfect results across all metrics. BB shows 100% performance across all metrics, while Deposits and Retail Products achieve 100% hit rates with HR@1 values exceeding 92%.

The model also performs exceptionally well for investment products, particularly Investments and Securities (100% hit rate, 99.26% HR@1), which contrasts with the Product Similarity model's weaker ranking performance for these products. This suggests that demographic and behavioral attributes captured by the Propensity Model are strongly predictive of investment product adoption.

Even for less common products, the Propensity Model generally outperforms other approaches. For Forex (51 test cases), it achieves a 96.08% hit rate, substantially higher than Product Similarity's 62.75%. For Cash Management (9 test cases), it achieves a perfect hit rate, though ranking accuracy remains a challenge (HR@1 of 11.11%).

The model's consistent performance across product categories demonstrates the effectiveness of the XGBoost algorithm in capturing complex relationships between customer attributes and product adoption patterns. The feature engineering and model optimization approaches detailed in the methodology section appear particularly effective for predicting individual customer preferences.

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
BB	0.991968	0.882642	0.910612	0.198394	0.795181	0.958501	2241
DEPOSITS	0.996861	0.886510	0.914688	0.199372	0.800448	0.962780	2230
RETAIL_PRODUCTS	0.375000	0.375000	0.375000	0.075000	0.375000	0.375000	40
REMITTANCE	1.000000	0.967111	0.975494	0.200000	0.941889	0.987893	413
INSURANCE	0.200000	0.056000	0.090948	0.040000	0.000000	0.120000	25
BUILDUP_F	0.942857	0.354286	0.501016	0.188571	0.057143	0.885714	35
CORPORATE_CARDS	0.750000	0.244444	0.368535	0.150000	0.000000	0.500000	12
FOREX	0.966667	0.821111	0.855945	0.193333	0.766667	0.833333	30
CASH_MANAGEMENT	1.000000	0.435288	0.579768	0.200000	0.006173	0.975309	162
CORPORATE_LOANS	1.000000	0.257576	0.436979	0.200000	0.000000	0.090909	66
INVESTMENTS_AND_SECURITIES	0.928571	0.803571	0.836304	0.185714	0.678571	0.928571	28
INVESTMENT_FUNDS	0.928571	0.595238	0.681113	0.185714	0.321429	0.928571	28
TRADE_SERVICES	0.777778	0.481481	0.558191	0.155556	0.222222	0.777778	9
Overall	0.984584	0.853867	0.887102	0.196916	0.757097	0.941154	5319

Table 4.4: Performance metrics for the Customer Similarity approach for Organizational Customers.

4.2.2 Performance for Organizational Customers

Customer Similarity Model Performance

For organizational customers, the Customer Similarity model (Table 4.4) shows a strong overall hit rate of 98.46%, with HR@1 of 75.71% and HR@3 of 94.12%. However, these aggregate metrics mask significant product-specific variations.

The model performs exceptionally well for the most common product categories: BB (99.20% hit rate, 79.52% HR@1) and Deposits (99.69% hit rate, 80.04% HR@1). For Remittance, it achieves perfect hit rate with extremely high ranking accuracy (94.19% HR@1), indicating strong customer similarity patterns for this service.

However, the model struggles significantly with certain product categories. For Insurance, the hit rate falls to just 20% with 0% HR@1, suggesting very weak customer similarity signals for organizational insurance products.

Similarly, for Retail Products, the hit rate is only 37.5%, though when the model does identify this product, it consistently ranks it at the top (HR@1 also 37.5%).

For Corporate Loans, while the model achieves a perfect hit rate, its ranking performance is poor (HR@1 of 0%, HR@3 of only 9.09%). This indicates that while the model recognizes corporate loans as relevant, it consistently ranks them too low in recommendation lists.

These results suggest that customer similarity captures different aspects of organizational purchasing patterns compared to individual customers, with stronger signals for transactional services (Remittance, BB, Deposits) than for specialized financial products (Insurance, Corporate Cards).

Product Similarity Model Performance

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
BB	1.0000	0.9931	0.9949	0.2000	0.9866	1.0000	2241
DEPOSITS	1.0000	1.0000	1.0000	0.2000	1.0000	1.0000	2230
RETAIL_PRODUCTS	0.8500	0.2767	0.4190	0.1700	0.0000	0.7000	40
REMITTANCE	1.0000	0.9988	0.9991	0.2000	0.9976	1.0000	413
INSURANCE	0.7600	0.2287	0.3574	0.1520	0.0000	0.3600	25
BUILDUP_F	0.8000	0.3081	0.4320	0.1600	0.0000	0.6857	35
CORPORATE_CARDS	0.5833	0.2153	0.3077	0.1167	0.0000	0.5000	12
FOREX	0.8667	0.4261	0.5391	0.1733	0.0667	0.8000	30
CASH_MANAGEMENT	0.9383	0.5239	0.6237	0.1877	0.3827	0.4074	162
CORPORATE_LOANS	1.0000	0.4033	0.5537	0.2000	0.0000	0.6970	66
INVESTMENTS_AND_SECURITIES	1.0000	0.8214	0.8682	0.2000	0.6429	1.0000	28
INVESTMENT_FUNDS	0.9286	0.7679	0.8099	0.1857	0.6071	0.9286	28
TRADE_SERVICES	0.8889	0.3500	0.4808	0.1778	0.1111	0.3333	9
Overall	0.9923	0.9532	0.9629	0.1985	0.9312	0.9671	5319

Table 4.5: Performance metrics for the Product Similarity approach for Organizational Customers.

The Product Similarity model (Table 4.5) demonstrates excellent overall performance for organizational customers, with a 99.23% hit rate, 93.12% HR@1, and 96.71% HR@3. These strong aggregate metrics are primarily driven by exceptional performance on the two most frequent products: BB and Deposits, which together represent approximately 84% of test cases.

For BB and Deposits, the model achieves perfect hit rates with near-perfect (BB, 98.66%) or perfect (Deposits, 100%) HR@1 scores. Similarly, for Remittance, it achieves a perfect hit rate with 99.76% HR@1, indicating extremely strong co-ownership patterns for these core banking products.

However, the model shows considerable weakness for several specialized products. For Retail Products, Insurance, Buildup F, and Corporate Cards, the model achieves moderate hit rates but 0% HR@1, indicating that while these products are eventually identified, they are consistently ranked too low. This pattern suggests limited co-ownership signals for these specialized products in the organizational context.

Investment products show mixed results: Investments and Securities achieves strong performance (100% hit rate, 64.29% HR@1), while Investment Funds performs similarly well (92.86% hit rate, 60.71% HR@1). This contrasts with the pattern seen for individual customers, where investment products showed weaker ranking performance.

Propensity Model Performance

The Propensity Model (Table 4.6) demonstrates strong overall performance for organizational customers, with a 99.19% hit rate, 92.95% HR@1, and 96.65% HR@3. These metrics are competitive with the Product Similarity model, though the patterns across individual products reveal important differences.

For high-frequency products, the model achieves excellent results: BB and Deposits both show perfect hit rates with HR@1 values exceeding 99.8%.

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
BB	1.0000	0.9993	0.9995	0.2000	0.9987	1.0000	2241
DEPOSITS	1.0000	0.9991	0.9993	0.2000	0.9982	1.0000	2230
RETAIL_PRODUCTS	0.8250	0.5925	0.6514	0.1650	0.4250	0.7500	40
REMITTANCE	1.0000	0.8909	0.9173	0.2000	0.8499	0.8717	413
INSURANCE	0.7600	0.3467	0.4475	0.1520	0.1600	0.4800	25
BUILDUP_F	0.8286	0.5281	0.6041	0.1657	0.3143	0.7429	35
CORPORATE_CARDS	0.5833	0.3056	0.3770	0.1167	0.0833	0.5833	12
FOREX	0.8667	0.7139	0.7518	0.1733	0.6333	0.7667	30
CASH_MANAGEMENT	0.9259	0.3137	0.4648	0.1852	0.0370	0.6111	162
CORPORATE_LOANS	1.0000	0.6869	0.7653	0.2000	0.5000	0.8485	66
INVESTMENTS_AND_SECURITIES	1.0000	1.0000	1.0000	0.2000	1.0000	1.0000	28
INVESTMENT_FUNDS	0.9286	0.6250	0.7045	0.1857	0.3214	0.9286	28
TRADE_SERVICES	0.8889	0.3500	0.4808	0.1778	0.1111	0.3333	9
Overall	0.9919	0.9506	0.9608	0.1984	0.9295	0.9665	5319

Table 4.6: Performance metrics for the Propensity Model approach for organizational customers.

For Investments and Securities, the model achieves perfect performance across all metrics, indicating that the XGBoost algorithm effectively captures the organizational attributes that predict adoption of this product.

For specialized business products, the Propensity Model often outperforms the other approaches on ranking metrics. For Retail Products, it achieves an HR@1 of 42.5% compared to 0% for Product Similarity, despite a slightly lower hit rate (82.5% vs. 85%). For Corporate Loans, it achieves an HR@1 of 50% compared to 0% for Product Similarity, with both models achieving 100% hit rates.

However, the model shows limitations for certain niche products. For Corporate Cards (12 test cases), it achieves only a 58.33% hit rate with 8.33% HR@1, and for Trade Services (9 test cases), it manages an 88.89% hit rate but only 11.11% HR@1. These results highlight the challenge of building effective prediction models for rarely purchased products, even with advanced machine learning techniques.

4.2.3 Comparative Analysis of Individual Models

After examining each model's performance independently, we now directly compare their effectiveness in generating accurate recommendations. We focus particularly on HR@1 and HR@3 metrics as they provide the most actionable insights into recommendation quality, reflecting a model's ability to place relevant products at the top of recommendation lists.

Model Comparison for Individual Customers

Product	Customer Similarity		Product Similarity		Propensity Model		Total Tested
	HR@1	HR@3	HR@1	HR@3	HR@1	HR@3	
INSURANCE	0.269939	0.720859	0.404908	0.957055	0.656442	0.984663	326
BB	0.336085	0.705189	0.995283	1.000000	1.000000	1.000000	848
DEPOSITS	0.485010	0.788969	0.998801	1.000000	0.997602	1.000000	834
RETAIL_PRODUCTS	0.294686	0.724499	0.979066	1.000000	0.921095	1.000000	621
INVESTMENTS_AND_SECURITIES	0.822222	0.918519	0.170370	0.866667	0.992593	1.000000	135
REMITTANCE	0.338028	0.684507	0.543662	1.000000	0.754930	0.994366	355
INVESTMENT_FUNDS	0.926230	0.983607	0.122951	0.909836	0.893443	0.983607	122
CORPORATE_LOANS	0.308962	0.801887	0.714623	0.997642	0.773585	0.974057	424
BUILDUP_F	0.217143	0.782857	0.165000	0.874286	0.428571	0.931429	175
FOREX	0.078431	0.607843	0.019608	0.274510	0.196078	0.588235	51
SECURITIES	0.588744	1.000000	0.102564	0.487179	0.205128	0.897436	39
CASH_MANAGEMENT	0.000000	0.444444	0.000000	0.000000	0.111111	0.444444	9
Overall	**0.381569**	**0.757044**	**0.757553**	**0.966489**	**0.862909**	**0.984259**	**3939**

Table 4.7: Comparison of HR@1 and HR@3 across all models for Individual Customers.

Table 4.7 reveals that for individual customers, the Propensity Model consistently outperforms the other approaches, achieving the highest overall HR@1 (86.29%) and HR@3 (98.43%) scores. This suggests that customer demographic and behavioral attributes, as captured by the machine learning model, are generally more predictive of product adoption than either similarity-based approach.

Despite the overall dominance of the Propensity Model, each approach demonstrates unique strengths for specific product categories:

- **Common Banking Products (BB, Deposits, Retail):** The Product Similarity and Propensity models both excel with these high-frequency products, achieving HR@1 values above 97% and nearly perfect HR@3 scores. The Customer Similarity model significantly underperforms for these products, with HR@1 values below 50%.

- **Investment Products:** A striking pattern emerges where Customer Similarity excels for Investment Funds (92.62% HR@1, 98.36% HR@3) and performs strongly for Investments and Securities (82.22% HR@1, 91.85% HR@3). Product Similarity shows surprisingly weak ranking for these same products (12.30% HR@1 for Investment Funds and 17.04% HR@1 for Investments and Securities). This suggests that investment product adoption is more strongly predicted by similar customer characteristics than by co-ownership patterns.

- **Securities:** Customer Similarity significantly outperforms other approaches, with an HR@1 of 58.87% compared to 20.51% for Propensity and just 10.26% for Product Similarity. This reinforces the pattern seen with other investment products, where customer-based approaches excel.

- **Low-Frequency Products:** All models struggle with products having few test cases. For Cash Management (9 test cases), only the Propensity Model achieves any success at HR@1 (11.11%), while all models perform relatively poorly for Forex (51 test cases) with HR@1 values below 20%.

The divergent performance across product categories highlights the complementary nature of these recommendation approaches. While the Propensity Model offers the strongest overall performance, there are specific product categories—particularly investment products—where Customer Similarity provides more accurate recommendations. This complementarity suggests potential benefits from an ensemble approach that leverages the strengths of each model for different product categories.

Model Comparison for Organizational Customers

Product	**Customer Similarity**		**Product Similarity**		**Propensity Model**		**Total Tested**
	HR@1	**HR@3**	**HR@1**	**HR@3**	**HR@1**	**HR@3**	
BB	0.795181	0.958501	0.9866	1.0000	0.9987	1.0000	2241
DEPOSITS	0.800448	0.962780	1.0000	1.0000	0.9982	1.0000	2230
RETAIL_PRODUCTS	0.375000	0.375000	0.0000	0.7000	0.4250	0.7500	40
REMITTANCE	0.941889	0.987893	0.9976	1.0000	0.8499	0.8717	413
INSURANCE	0.000000	0.120000	0.0000	0.3600	0.1600	0.4800	25
BUILDUP_F	0.057143	0.885714	0.0000	0.6857	0.3143	0.7429	35
CORPORATE_CARDS	0.000000	0.500000	0.0000	0.5000	0.0833	0.5833	12
FOREX	0.766667	0.833333	0.0667	0.8000	0.6333	0.7667	30
CASH_MANAGEMENT	0.006173	0.975309	0.3827	0.4074	0.0370	0.6111	162
CORPORATE_LOANS	0.000000	0.090909	0.0000	0.6970	0.5000	0.8485	66
INVESTMENTS_AND_SECURITIES	0.678571	0.928571	0.6429	1.0000	1.0000	1.0000	28
INVESTMENT_FUNDS	0.321429	0.928571	0.6071	0.9286	0.3214	0.9286	28
TRADE_SERVICES	0.222222	0.777778	0.1111	0.3333	0.1111	0.3333	9
Overall	**0.757097**	**0.941154**	**0.9312**	**0.9671**	**0.9295**	**0.9665**	**5319**

Table 4.8: Comparison of HR@1 and HR@3 across all models for Organizational Customers.

Table 4.8 reveals that for organizational customers, the performance comparison is more nuanced than for individual customers. The Propensity Model (92.95%) and Product Similarity (93.12%) achieve similar overall HR@1 scores, both substantially outperforming Customer Similarity (75.71%). However, for HR@3, all three models perform relatively well, with the Customer Similarity (94.12%) competitive with the Propensity Model (96.65%) and Product Similarity (96.71%).

Several product-specific patterns emerge from this comparison:

- **Core Banking Products (BB, Deposits):** All three models perform strongly for these most frequently tested products, with the Propensity Model and Product Similarity achieving near-perfect HR@1 scores (>98.6%). Customer Similarity performs well but lags somewhat with HR@1 scores around 80%.

- **Remittance:** Product Similarity demonstrates exceptional performance (99.76% HR@1), outperforming both Customer Similarity (94.19%) and the Propensity Model (84.99%). This suggests strong co-ownership patterns for remittance services among organizational customers.

- **Specialized Business Products:** For Corporate Loans, only the Propensity Model achieves meaningful ranking accuracy (50% HR@1), while the other models show 0% HR@1. This demonstrates the value of machine learning for capturing complex patterns in specialized product adoption.

- **Cash Management:** Product Similarity significantly outperforms other approaches at HR@1 (38.27% vs. 3.70% for Propensity and 0.62% for Customer Similarity), while Customer Similarity excels at HR@3 (97.53%). This suggests that different recommendation approaches capture different aspects of cash management service adoption.

- **Investments and Securities:** The Propensity Model achieves perfect scores (100% HR@1 and HR@3), outperforming both Customer Similarity

(67.86% HR@1) and Product Similarity (64.29% HR@1). This contrasts with the pattern seen for individual customers, where Customer Similarity showed particular strength for investment products.

- **Low-Frequency Products:** All models struggle with products that have few test cases, particularly Insurance (25 cases) and Corporate Cards (12 cases), with HR@1 values below 16% across all models.

For organizational customers, the complementary strengths of different models are even more pronounced than for individual customers. While the Propensity Model holds a slight edge in overall HR@1 performance, there are multiple product categories where either Customer Similarity (Forex, Trade Services) or Product Similarity (Remittance, Cash Management, Investment Funds) provides more accurate recommendations.

The stark differences in model performance across product categories suggest that organizational customer purchasing patterns are more complex and varied than individual customer patterns. This complexity creates a particularly compelling case for an ensemble approach that can leverage the unique strengths of each model for different product categories.

4.3 Ensemble Model Performance

After establishing the performance of individual recommendation models, we now evaluate the effectiveness of our ensemble approaches. The key question is whether combining these models with appropriate weighting can leverage their complementary strengths to achieve better performance than any single model alone.

4.3.1 Per-Model Weight Optimization

The per-model weight optimization approach involves finding the optimal combination of weights for the three recommendation models that maximizes performance for a specific product category. This allows us to understand which models contribute most effectively to each product's recommendation performance.

Weight Optimization for Individual Customers

Product Category	Customer Similarity	Product Similarity	Propensity
INSURANCE	0.00	0.20	0.80
BB	0.10	0.70	0.20
DEPOSITS	0.30	0.60	0.10
RETAIL_PRODUCTS	0.10	0.60	0.30
INVESTMENTS_AND_SECURITIES	0.20	0.00	0.80
REMITTANCE	0.00	0.20	0.80
INVESTMENT_FUNDS	0.90	0.00	0.10
CORPORATE_LOANS	0.40	0.00	0.60
BUILDUP_F	0.60	0.00	0.40
FOREX	0.30	0.00	0.70
SECURITIES	0.80	0.10	0.10
CASH_MANAGEMENT	0.00	0.00	1.00

Table 4.9: Product-Specific Optimized Model Weights for Hitrate@1 for Individual Customers

Table 4.9 presents the optimal weights for each product category that would maximize the HR@1 for that specific product. These weights reveal several important patterns:

- **Product-Specific Optimization:** The optimal weights vary substantially across product categories, confirming that different recommendation approaches excel for different products.

- **Propensity Model Dominance:** For several products (Insurance, Investments and Securities, Remittance, Forex, Cash Management), the Propensity Model receives the highest weight (0.70-1.00), reflecting its strong overall performance.

- **Customer Similarity for Investment Products:** For Investment Funds (0.90) and Securities (0.80), Customer Similarity receives the highest weight, consistent with our earlier finding that this model excels for investment products.

- **Product Similarity for Common Banking Products:** For high-frequency products like BB, Deposits, and Retail Products, the Product Similarity model receives the highest weight (0.60-0.70), despite the Propensity Model's strong individual performance for these categories.

- **Zero Weights:** Several products have zero weights assigned to certain models, effectively removing them from the ensemble for those categories. This indicates that these models may introduce noise rather than useful signals for these specific products.

To illustrate the impact of using product-specific optimized weights, we can examine the performance when applying the weights that are optimized for a specific product, such as Securities.

Table 4.10 shows the performance when applying the weight configuration optimized for Securities (Customer Similarity: 0.80, Product Similarity: 0.10, Propensity: 0.10). As expected, this configuration significantly improves the HR@1 for Securities to 61.54%, a substantial increase from the Propensity Model's 20.51

- **Positive Transfer:** Investment Funds achieves 98.36% HR@1, outper-

Product	Hit Rate	MRR	NCDG	Precision	HR@1	HR@3
INSURANCE	0.996933	0.673006	0.755672	0.199387	0.429448	0.957055
BB	1.000000	0.921384	0.941672	0.200000	0.854953	0.997642
DEPOSITS	1.000000	0.913869	0.936224	0.200000	0.835731	0.997602
RETAIL_PRODUCTS	1.000000	0.887815	0.916874	0.200000	0.789050	1.000000
INVESTMENTS_AND_SECURITIES	1.000000	0.931975	0.948811	0.200000	0.896296	0.962963
REMITTANCE	1.000000	0.688920	0.767513	0.200000	0.478873	0.901408
INVESTMENT_FUNDS	0.991803	0.985246	0.986777	0.198361	0.983607	0.983607
CORPORATE_LOANS	0.992925	0.835810	0.875108	0.198585	0.738208	0.919811
BUILDUP_F	0.960000	0.532476	0.640784	0.192000	0.222857	0.794286
FOREX	0.725490	0.319608	0.420407	0.145098	0.098039	0.529412
SECURITIES	1.000000	0.788462	0.842844	0.200000	0.615385	0.974359
CASH_MANAGEMENT	0.666667	0.198148	0.312201	0.133333	0.000000	0.222222
Overall	**0.992638**	**0.838085**	**0.877462**	**0.160934**	**0.722011**	**0.958872**

Table 4.10: Performance metrics using Securities-optimized weights for Individual Customers

forming the Propensity Model's 89.34

- **Performance Deterioration:** For many other products, performance declines compared to the Propensity Model. For example, BB achieves 85.50% HR@1 (vs. Propensity's 100%), and Insurance achieves 42.94% (vs. Propensity's 65.64%).
- **Overall Impact:** When using Securities-optimized weights, the overall HR@1 drops to 72.20% from the Propensity Model's 86.29

These results demonstrate that while product-specific weight optimization can significantly improve performance for targeted products, applying these weights universally across all products often leads to lower overall performance. This highlights the need for a more sophisticated approach that can apply different weights for different products—precisely the goal of the greedy weight matrix optimization method.

Weight Optimization for Organizational Customers

Product Category	Customer Similarity	Product Similarity	Propensity
BB	0.30	0.50	0.20
DEPOSITS	0.10	0.90	0.00
RETAIL_PRODUCTS	0.00	0.20	0.80
REMITTANCE	0.10	0.50	0.40
INSURANCE	0.10	0.20	0.70
BUILDUP_F	0.30	0.00	0.70
CORPORATE_CARDS	0.60	0.00	0.40
FOREX	1.00	0.00	0.00
CASH_MANAGEMENT	0.10	0.90	0.00
CORPORATE_LOANS	0.00	0.00	1.00
INVESTMENTS_AND_SECURITIES	0.10	0.70	0.20
INVESTMENT_FUNDS	0.70	0.30	0.00
TRADE_SERVICES	1.00	0.00	0.00

Table 4.11: Product-Specific Optimized Model Weights for Hitrate@1 for Organizational Customers

For organizational customers, Table 4.11 shows the optimal weights for each product category. These weights reveal different patterns compared to individual customers:

- **Greater Role for Customer Similarity:** For several products (Corporate Cards, Forex, Investment Funds, Trade Services), Customer Similarity receives the highest weight (0.60-1.00), suggesting this model captures important aspects of organizational purchasing behavior.
- **Product Similarity Dominance for Core Banking:** For BB, Deposits, Remittance, and Cash Management, Product Similarity receives the highest weight (0.50-0.90), reflecting the strong co-ownership patterns for these products among organizational customers.
- **Propensity Model for Specialized Products:** For Retail Products, Insurance, Buildup F, and Corporate Loans, the Propensity Model re-

ceives the highest weight (0.70-1.00), highlighting its effectiveness for these more specialized offerings.

- **Complete Specialization:** For Forex and Trade Services, Customer Similarity receives a weight of 1.00, while for Corporate Loans, Propensity receives a weight of 1.00, indicating that a single model completely dominates for these products.

To illustrate the impact of product-specific weights for organizational customers, we examine the performance when applying the weights optimized for Cash Management.

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3
BB	0.9946	0.9906	0.9917	0.1989	0.9866	0.9946
DEPOSITS	1.0000	1.0000	1.0000	0.2000	1.0000	1.0000
RETAIL_PRODUCTS	0.4000	0.1333	0.2000	0.0800	0.0000	0.4000
REMITTANCE	1.0000	0.9988	0.9991	0.2000	0.9976	1.0000
INSURANCE	0.2400	0.0680	0.1096	0.0480	0.0000	0.0400
BUILDUP_F	0.9429	0.3090	0.4667	0.1886	0.0000	0.8571
CORPORATE_CARDS	0.7500	0.2306	0.3570	0.1500	0.0000	0.3333
FOREX	0.9667	0.4822	0.6055	0.1933	0.1000	0.8333
CASH_MANAGEMENT	1.0000	0.6851	0.7671	0.2000	0.3827	0.9815
CORPORATE_LOANS	0.9848	0.2508	0.4278	0.1970	0.0000	0.0909
INVESTMENTS_AND_SECURITIES	1.0000	0.8214	0.8682	0.2000	0.6429	1.0000
INVESTMENT_FUNDS	0.9286	0.7679	0.8099	0.1857	0.6071	0.9286
TRADE_SERVICES	0.7778	0.3556	0.4610	0.1556	0.1111	0.6667

Table 4.12: Performance metrics using Cash Management-optimized weights for Organizational Customers

Table 4.12 shows the performance when applying the weight configuration optimized for Cash Management (Customer Similarity: 0.10, Product Similarity: 0.90, Propensity: 0.00). This configuration successfully improves the HR@1 for Cash Management to 38.27%, substantially outperforming the Propensity Model's 3.70%. However, this specialization has varying effects on other products:

- **Core Banking Products:** BB, Deposits, and Remittance maintain excellent performance with HR@1 values of 98.66%, 100%, and 99.76% respectively, comparable to or better than the Propensity Model.
- **Specialized Products:** Several products (Retail Products, Insurance, Buildup F, Corporate Cards, Corporate Loans) show 0% HR@1 with these weights, indicating that the Cash Management-optimized configuration performs poorly for these products.
- **Investment Products:** Investments and Securities (64.29%) and Investment Funds (60.71%) maintain moderate performance, though Investments and Securities significantly underperforms the Propensity Model's perfect 100%.

These results from both customer types highlight a fundamental challenge in recommendation system design: optimizing for one product often comes at the expense of others. This trade-off underscores the value of our greedy weight matrix optimization approach, which allows for product-specific weight configurations rather than applying a single weight configuration across all products.

4.3.2 Implications for Ensemble Design

The per-model weight optimization analysis reveals several important insights for designing an effective ensemble recommendation system:

- **Model Complementarity:** The substantial variation in optimal weights across products confirms that our three models capture complementary signals. For example, Customer Similarity excels for investment products, Product Similarity for core banking products, and the Propensity Model for specialized offerings.

- **Product-Specific Optimization:** The dramatic differences in optimal weights across products suggest that a one-size-fits-all approach is suboptimal. An effective ensemble should apply different weights for different product categories.

- **Weight Matrix Potential:** The ability to improve specific product performance through targeted weight optimization demonstrates the potential value of a weight matrix approach that can apply product-specific weights.

- **Specialization vs. Generalization:** The trade-offs observed when applying product-specific weights highlight the tension between specializing for certain products versus maintaining strong general performance.

These findings motivate our greedy weight matrix optimization approach, which we analyze in the next section. By allowing for product-specific weight configurations, this more advanced ensemble method aims to capture the benefits of specialization while avoiding the performance penalties observed when applying a single weight configuration across all products.

4.3.3 Greedy Weight Matrix Optimization

Building on the insights from the per-model weight optimization, we now examine our more advanced ensemble approach: greedy weight matrix optimization. This approach creates a product-specific weight matrix, allowing each product to have its own combination of model weights, while optimizing for the overall Hit Rate@3 performance across all products.

Optimization Approach for Individual Customers

Product	Customer Similarity	Product Similarity	Propensity Model
INSURANCE	0.2000	0.1000	0.7000
BB	0.3333	0.3333	0.3333
DEPOSITS	0.3333	0.3333	0.3333
RETAIL_PRODUCTS	0.3333	0.3333	0.3333
INVESTMENTS_AND_SECURITIES	0.3333	0.3333	0.3333
REMITTANCE	0.5000	0.5000	0.0000
INVESTMENT_FUNDS	0.3333	0.3333	0.3333
CORPORATE_LOANS	0.3000	0.5000	0.2000
BUILDUP_F	0.0000	0.0000	1.0000
FOREX	1.0000	0.0000	0.0000
SECURITIES	1.0000	0.0000	0.0000
CASH_MANAGEMENT	1.0000	0.0000	0.0000

Table 4.13: Optimized Model Weights from Greedy Search for Individual Customers

Table 4.13 presents the optimized weights derived from the greedy search for individual customers, targeting overall HR@3 performance. Several notable patterns emerge:

- **Selective Model Exclusion:** For some products, the greedy search entirely excludes certain models by assigning zero weights. For example, the Propensity Model is excluded for Remittance, while both Product Similarity and Propensity are excluded for Forex, Securities, and Cash Management.
- **Customer Similarity Dominance:** For specialized products with fewer samples (Forex, Securities, Cash Management), Customer Similarity re-

ceives a weight of 1.0, effectively using only this model. This aligns with our earlier observation that Customer Similarity can excel for certain niche products.

- **Propensity Model Role:** The Propensity Model receives the highest weight (0.7) for Insurance and exclusive weight (1.0) for Buildup F, reflecting its strength for these categories.

The varied weight distributions across products confirm the value of the product-specific optimization approach, even when optimizing for overall HR@3 performance rather than individual product metrics.

Product	Hit Rate	MRR	NDCG	HR@1	HR@3
INSURANCE	1.000000	0.811500	0.859962	0.656442	0.987730
BB	1.000000	0.998821	0.999130	0.997642	1.000000
DEPOSITS	1.000000	0.998801	0.999115	0.997602	1.000000
RETAIL_PRODUCTS	1.000000	0.991143	0.994463	0.982287	1.000000
INVESTMENTS_AND_SECURITIES	1.000000	0.969136	0.977159	0.940741	1.000000
REMITTANCE	1.000000	0.711668	0.786596	0.467606	1.000000
INVESTMENT_FUNDS	1.000000	0.963115	0.972516	0.934426	0.983607
CORPORATE_LOANS	1.000000	0.842374	0.883501	0.691038	1.000000
BUILDUP_F	0.994286	0.667610	0.770963	0.428571	0.931429
FOREX	0.862745	0.343791	0.477765	0.078431	0.607843
SECURITIES	1.000000	0.773504	0.831801	0.589744	1.000000
CASH_MANAGEMENT	0.666667	0.222222	0.332476	0.000000	0.444444

Table 4.14: Ensemble System Performance for Individual Customers

Table 4.14 shows the performance achieved using the greedy weight matrix optimization for individual customers. The results reveal several important insights:

- **Enhanced HR@3 Performance:** The ensemble achieves perfect or near-perfect HR@3 for 8 out of 12 products, with an overall HR@3 performance exceeding the best individual model. This confirms the effectiveness of optimizing specifically for this metric.

- **Improved Product Coverage:** The ensemble achieves a 100% hit rate for 10 out of 12 products, compared to 9 out of 12 for the Propensity Model alone. This indicates improved coverage of relevant products.
- **Collateral Improvements in HR@1:** Despite optimizing for HR@3, the ensemble also achieves higher HR@1 than the Propensity Model for several products. For example, Retail Products (98.23% vs. 92.11%), Securities (58.97% vs. 20.51%), and Investment Funds (93.44% vs. 89.34%).
- **Consistent Performance:** For products where the Propensity Model already excelled (BB, Deposits), the ensemble maintains this high performance with HR@1 values exceeding 99.7%.
- **Remittance Trade-off:** For Remittance, the ensemble achieves 46.76% HR@1 compared to the Propensity Model's 75.49%, illustrating a trade-off where optimizing for HR@3 sometimes reduces HR@1 performance for specific categories.

The greedy weight matrix approach demonstrates significant improvements over the simpler per-model weight optimization, successfully leveraging the strengths of different models for different product categories while optimizing for overall HR@3 performance.

Optimization Approach for Organizational Customers

Table 4.15 presents the optimized weights for organizational customers, explicitly optimized for overall HR@3 performance. Similar to individual customers,

Product	Propensity	Product Similarity	Customer Similarity
BB	0.33	0.33	0.33
BUILDUP_F	0.00	0.00	1.00
CASH_MANAGEMENT	0.00	0.40	0.60
CORPORATE_CARDS	0.20	0.50	0.30
CORPORATE_LOANS	0.70	0.00	0.30
DEPOSITS	0.33	0.33	0.33
FOREX	0.20	0.70	0.10
INSURANCE	1.00	0.00	0.00
INVESTMENTS_AND_SECURITIES	0.33	0.33	0.33
INVESTMENT_FUNDS	0.33	0.33	0.33
REMITTANCE	0.33	0.33	0.33
RETAIL_PRODUCTS	1.00	0.00	0.00
TRADE_SERVICES	0.00	0.00	1.00

Table 4.15: Optimized Model Weights from Greedy Search for Organizational Customers (HitRate@3)

we observe varied weight distributions across products:

- **Balanced Approach:** For five products (BB, Deposits, Investments and Securities, Investment Funds, Remittance), the greedy approach converged to equal weights (0.33) for each model, suggesting that a balanced ensemble works best for these products within the global optimization framework.

- **Model Specialization:** For certain products, the approach relies exclusively on a single model. For Insurance and Retail Products, it uses only the Propensity Model (weight 1.0), while for Buildup F and Trade Services, it relies entirely on Customer Similarity (weight 1.0).

- **Customer Similarity Prominence:** Customer Similarity receives substantial weights for several products, particularly Buildup F (1.0), Cash Management (0.6), and Trade Services (1.0), confirming its value for certain organizational products.

- **Product Similarity Role:** Product Similarity dominates for Forex (0.7) and Corporate Cards (0.5), reflecting its strength in capturing co-ownership patterns for these products among organizational customers.

Product	Hit Rate	MRR	NDCG	HR@1	HR@3
BB	1.000000	0.999331	0.999506	0.998661	1.000000
DEPOSITS	1.000000	0.999327	0.999503	0.998655	1.000000
REMITTANCE	1.000000	0.996368	0.997319	0.992736	1.000000
CASH_MANAGEMENT	1.000000	0.652675	0.743151	0.320988	0.987654
CORPORATE_LOANS	1.000000	0.638131	0.729922	0.378788	0.878788
INVESTMENTS_AND_SECURITIES	1.000000	0.982143	0.986819	0.964286	1.000000
BUILDUP_F	0.942857	0.354286	0.501016	0.057143	0.885714
INVESTMENT_FUNDS	0.928571	0.928571	0.928571	0.928571	0.928571
FOREX	0.866667	0.450000	0.558317	0.066667	0.866667
CORPORATE_CARDS	0.833333	0.304167	0.435675	0.000000	0.583333
RETAIL_PRODUCTS	0.825000	0.592500	0.651438	0.425000	0.750000
TRADE_SERVICES	0.777778	0.481481	0.558191	0.222222	0.777778
INSURANCE	0.760000	0.346667	0.447536	0.160000	0.480000

Table 4.16: Ensemble System Performance for Organizational Customers

Table 4.16 shows the performance of the greedy weight matrix ensemble for organizational customers, optimized for overall HR@3. Several key observations can be made:

- **Strong HR@3 Performance:** The ensemble achieves HR@3 exceeding 87% for 9 out of 13 products, with perfect scores for core banking products, demonstrating the effectiveness of the optimization target.
- **Near-Perfect Core Banking:** For BB, Deposits, and Remittance, the ensemble achieves exceptional performance with HR@1 exceeding 99.2%

and perfect HR@3, comparable to or better than the best individual model.

- **Improved Investment Performance:** For Investments and Securities, the ensemble achieves 96.43% HR@1 and 100% HR@3, slightly lower than the Propensity Model's perfect 100% for HR@1 but maintaining perfect HR@3. For Investment Funds, it achieves 92.86% HR@1 and HR@3, substantially outperforming the Propensity Model's 32.14% HR@1.
- **Cash Management Enhancement:** For Cash Management, the ensemble achieves 32.10% HR@1 and 98.77% HR@3, significantly outperforming the Propensity Model's 3.70% HR@1 and 61.11% HR@3.
- **Mixed Results for Specialized Products:** For Corporate Loans, the ensemble achieves 37.88% HR@1 and 87.88% HR@3, with an HR@1 below the Propensity Model's 50.0% but an HR@3 above the Propensity Model's 84.85%.

The greedy weight matrix approach optimized for HR@3 for organizational customers demonstrates effective balancing of model strengths across different product categories, with particularly notable improvements in HR@3 for multiple products.

4.3.4 Comparative Analysis: Greedy Ensemble vs. Best Individual Model

To assess the value added by the greedy weight matrix ensemble optimized for HR@3, we directly compare its performance against the best individual model—the Propensity Model—for both customer types.

Table 4.17 directly compares the HR@3 performance of the greedy ensemble and the Propensity Model. This comparison highlights the effectiveness of the ensemble approach when optimizing specifically for HR@3:

Product	Individual Customers			Organizational Customers		
	Propensity HR@3	Greedy HR@3	Difference	Propensity HR@3	Greedy HR@3	Difference
BB	1.000000	1.000000	0.000000	1.0000	1.000000	0.000000
DEPOSITS	1.000000	1.000000	0.000000	1.0000	1.000000	0.000000
RETAIL_PRODUCTS	1.000000	1.000000	0.000000	0.7500	0.750000	0.000000
INVESTMENTS_AND_SECURITIES	1.000000	1.000000	0.000000	1.0000	1.000000	0.000000
REMITTANCE	0.994366	1.000000	+0.005634	0.8717	1.000000	+0.128300
INVESTMENT_FUNDS	0.983607	0.983607	0.000000	0.9286	0.928571	0.000000
CORPORATE_LOANS	0.974057	1.000000	+0.025943	0.8485	0.878788	+0.030288
BUILDUP_F	0.931429	0.931429	0.000000	0.7429	0.885714	+0.142814
FOREX	0.588235	0.607843	+0.019608	0.7667	0.866667	+0.099967
SECURITIES	0.897436	1.000000	+0.102564	-	-	-
CASH_MANAGEMENT	0.444444	0.444444	0.000000	0.6111	0.987654	+0.376554
INSURANCE	0.984663	0.987730	+0.003067	0.4800	0.480000	0.000000
CORPORATE_CARDS	-	-	-	0.5833	0.583333	0.000000
TRADE_SERVICES	-	-	-	0.3333	0.777778	+0.444478

Table 4.17: Comparison of HR@3 between Propensity Model and Greedy Weight Matrix Ensemble

- **Individual Customers:** The greedy ensemble achieves better or equal HR@3 for all products compared to the Propensity Model. Notable improvements include Securities (+10.26 percentage points), Corporate Loans (+2.59 pp), and Forex (+1.96 pp).
- **Organizational Customers:** More substantial improvements are observed, with dramatic gains for Trade Services (+44.45 pp), Cash Management (+37.66 pp), Buildup F (+14.28 pp), and Remittance (+12.83 pp).
- **Consistent Performance:** For products where the Propensity Model already achieved perfect or near-perfect HR@3, the ensemble maintains this high performance.
- **No Performance Degradation:** Importantly, there are no cases where the ensemble performs worse than the Propensity Model for HR@3, confirming the effectiveness of the optimization approach.

While the greedy approach was optimized for HR@3, Table 4.18 examines

Product	Individual Customers			Organizational Customers		
	Propensity HR@1	Greedy HR@1	Difference	Propensity HR@1	Greedy HR@1	Difference
BB	1.000000	0.997642	-0.002358	0.9987	0.998661	-0.000039
DEPOSITS	0.997602	0.997602	0.000000	0.9982	0.998655	+0.000455
RETAIL_PRODUCTS	0.921095	0.982287	+0.061192	0.4250	0.425000	0.000000
INVESTMENTS_AND_SECURITIES	0.992593	0.940741	-0.051852	1.0000	0.964286	-0.035714
REMITTANCE	0.754930	0.467606	-0.287324	0.8499	0.992736	+0.142836
INVESTMENT_FUNDS	0.893443	0.934426	+0.040983	0.3214	0.928571	+0.607171
CORPORATE_LOANS	0.773585	0.691038	-0.082547	0.5000	0.378788	-0.121212
BUILDUP_F	0.428571	0.428571	0.000000	0.3143	0.057143	-0.257157
FOREX	0.196078	0.078431	-0.117647	0.6333	0.066667	-0.566633
SECURITIES	0.205128	0.589744	+0.384616	-	-	-
CASH_MANAGEMENT	0.111111	0.000000	-0.111111	0.0370	0.320988	+0.283988
INSURANCE	0.656442	0.656442	0.000000	0.1600	0.160000	0.000000
CORPORATE_CARDS	-	-	-	0.0833	0.000000	-0.083300
TRADE_SERVICES	-	-	-	0.1111	0.222222	+0.111122

Table 4.18: Comparison of HR@1 between Propensity Model and Greedy Weight Matrix Ensemble

its impact on HR@1 performance. As expected, the results are more mixed:

- **Significant Improvements for Key Products:** For individual customers, the greedy ensemble substantially improves HR@1 for Retail Products (+6.12 percentage points), Investment Funds (+4.10 pp), and Securities (+38.46 pp). For organizational customers, dramatic improvements are seen for Investment Funds (+60.72 pp), Cash Management (+28.40 pp), Remittance (+14.28 pp), and Trade Services (+11.11 pp).
- **Core Banking Stability:** For high-frequency products like BB and Deposits, the ensemble maintains the excellent performance of the Propensity Model for both customer types, with differences less than 0.25 percentage points.
- **Performance Trade-offs:** For some products, the HR@1 performance decreases when optimizing for HR@3, with notable decreases for Remittance (-28.73 pp) and Forex (-11.76 pp) for individual customers, and

Forex (-56.66 pp) and Buildup F (-25.72 pp) for organizational customers.

These results demonstrate the inherent trade-offs when optimizing for HR@3 instead of HR@1, though many products still see impressive gains in HR@1 despite this not being the primary optimization target.

4.3.5 Advantages of Greedy Weight Matrix Optimization

The greedy weight matrix optimization approach, specifically targeting overall HR@3 performance, offers several key advantages:

- **Comprehensive Recommendation Quality:** By optimizing for HR@3, the approach ensures that correct recommendations consistently appear within the top three positions, providing customers with highly relevant options even when the exact top recommendation is not perfect.
- **Product-Specific Customization:** By allowing each product to have its own optimized weight configuration, the approach can tailor recommendations to the unique patterns of each product category while still targeting overall performance.
- **Model Complementarity:** The approach effectively leverages the complementary strengths of different models, using Customer Similarity for investment products, Product Similarity for core banking, and the Propensity Model for specialized offerings.
- **Robust Performance:** The greedy approach achieves consistently strong HR@3 performance across diverse product categories, demonstrating its robustness in handling different recommendation scenarios.
- **Implementation Efficiency:** Despite its sophistication, the approach remains computationally efficient, requiring only a simple weighted sum

of normalized scores for each product.

These advantages highlight the value of the greedy weight matrix optimization as a practical and effective approach for enhancing recommendation quality in a banking context, particularly when the goal is to ensure that relevant products consistently appear within the top three recommendations.

4.3.6 Additional Results: Temporal Validation

To further assess the models' robustness and simulate a more dynamic scenario, we conducted an additional validation incorporating a temporal aspect. In this approach, instead of hiding a randomly selected owned product, we specifically identified and withheld the most recently acquired product for each customer, provided it was present in the sparse matrix representing product ownership. This tests the models' ability to predict the immediate next product a customer is likely to acquire based on their history. The performance metrics under this temporal validation scheme are presented below for both individual and organizational customers.

Temporal Validation for Individual Customers

Temporal Validation for Organizational Customers

Comparing the results from the temporal validation (Tables 4.19 and 4.20) with the standard leave-one-out evaluation (Tables A.27 and A.28) reveals a significant decrease in performance across most models and customer types when predicting the most recently acquired product. For individual customers, the best overall HR@1 drops from 86.3% (Propensity Model, Table A.27) to 71.6% (Propensity Model, Table 4.19). A similar, more pronounced drop is observed for organizational customers, where the best overall HR@1 falls from 93.1%

Product	Customer Similarity		Product Similarity		Propensity Model		Total Tested
	HR@1	HR@3	HR@1	HR@3	HR@1	HR@3	
INSURANCE	0.182927	0.689024	0.304878	0.926829	0.542683	0.969512	164
RETAIL_PRODUCTS	0.333333	0.787302	0.984127	1.000000	0.930159	1.000000	315
BUILDUP_F	0.283019	0.792453	0.113208	0.716981	0.358491	0.849057	53
REMITTANCE	0.333333	0.771429	0.380952	1.000000	0.761905	1.000000	105
FOREX	0.035714	0.464286	0.035714	0.214286	0.142857	0.428571	28
CASH_MANAGEMENT	0.000000	0.400000	0.000000	0.000000	0.000000	0.400000	5
SECURITIES	0.285714	1.000000	0.000000	0.000000	0.000000	0.857143	7
Overall	**0.277696**	**0.747415**	**0.601182**	**0.909897**	**0.716396**	**0.951255**	**677**

Table 4.19: Temporal Validation: Comparison of HR@1 and HR@3 across models for Individual Customers.

Product	Customer Similarity		Product Similarity		Propensity Model		Total Tested
	HR@1	HR@3	HR@1	HR@3	HR@1	HR@3	
RETAIL_PRODUCTS	0.085714	0.428571	0.000000	0.785714	0.500000	0.785714	28
REMITTANCE	0.996479	0.996479	1.000000	1.000000	0.867547	0.899394	284
INSURANCE	0.040000	0.200000	0.000000	0.200000	0.466667	0.000000	15
BUILDUP_F	0.064516	0.870968	0.000000	0.677419	0.354839	0.741935	31
CASH_MANAGEMENT	0.000000	0.080915	0.366197	0.380282	0.007042	0.605634	142
CORPORATE_LOANS	0.000000	0.090909	0.000000	0.672727	0.436364	0.818182	55
FOREX	0.000000	0.782609	0.000000	0.782609	0.652174	0.782609	23
INVESTMENT_FUNDS	0.111111	0.777778	0.666667	0.777778	0.111111	0.777778	9
TRADE_SERVICES	0.333333	0.666667	0.166667	0.500000	0.166667	0.500000	6
Overall	**0.635746**	**0.787149**	**0.617873**	**0.804579**	**0.663959**	**0.775923**	**593**

Table 4.20: Temporal Validation: Comparison of HR@1 and HR@3 across models for Organizational Customers.

(Product Similarity, Table A.28) to 66.4% (Propensity Model, Table 4.20). This marked decline highlights the challenge of predicting the immediate next purchase and underscores the importance of developing models that are explicitly temporal-aware. The current models, while effective in identifying generally relevant products based on overall ownership patterns, struggle to capture the specific sequence of product adoption. Incorporating temporal dynamics, such as the order and timing of product acquisitions, is a crucial direction for future work to improve the predictive power and real-world applicability of the NBO system.

CHAPTER V

Conclusions

This research has developed and evaluated a comprehensive recommendation system for banking products that integrates multiple recommendation approaches through a matrix-based ensemble methodology. The work addresses the significant challenge of identifying the most appropriate products for banking customers, a task made complex by the diverse product offerings and varying customer preferences across both individual and organizational segments in the financial services sector.

5.1 Summary of Research Contributions

Our comparative analysis of the three recommendation approaches revealed that propensity models generally outperform similarity-based approaches across most product categories. For individual customers, the Propensity Model achieved superior performance (86.29% HR@1, 98.43% HR@3) compared to Product Similarity (75.76% HR@1, 96.65% HR@3) and Customer Similarity (38.16% HR@1, 75.70% HR@3). For organizational customers, both the Propensity Model (92.95% HR@1, 96.65% HR@3) and Product Similarity (93.12% HR@1, 96.71% HR@3) demonstrated strong overall performance. For high-prevalence products like BB, DEPOSITS, and RETAIL_PRODUCTS, the propensity model achieved near-perfect hit rates with exceptional ranking quality across both customer segments.

While propensity models demonstrated overall dominance, our analysis uncovered specific cases where alternative approaches performed better. For individual customers, the Customer Similarity model significantly outperformed other approaches for investment products, achieving 92.62% HR@1 for INVESTMENT_FUNDS (compared to 89.34% for Propensity) and 82.22% HR@1 for INVESTMENTS_AND_SECURITIES (compared to 17.04% for Product Similarity). For organizational customers, Product Similarity excelled for REMITTANCE (99.76% HR@1 vs. 84.99% for Propensity) and CASH_MANAGEMENT (38.27% HR@1 vs. 3.70% for Propensity). These isolated but significant exceptions suggested that complementary recommendation signals exist across different approaches, providing a clear rationale for our ensemble methodology.

A critical finding of this research was the substantial impact of product prevalence on recommendation performance. All models performed significantly better for products with high positivity rates (common products that many customers already own) compared to those with low positivity rates. For example, the Propensity Model achieved perfect or near-perfect performance for BB and DEPOSITS across both customer types, while struggling with CASH_MANAGEMENT (11.11% HR@1 for individual customers, 3.70% HR@1 for organizational customers). This pattern raises important questions about real-world applicability, as the system demonstrates stronger performance for products customers are more likely to already possess, while struggling with niche products that might represent valuable cross-selling opportunities. The extreme class imbalance for some products (e.g., CASH_MANAGEMENT with only 9 test cases for individual customers) presents a fundamental challenge that may not be fully resolvable through methodological improvements alone.

The matrix-based weight optimization methodology revealed two effective approaches to ensemble modeling. The per-model weight optimization demonstrated that weights optimized for specific products could significantly enhance performance for those targeted products—for example, Securities-optimized weights improved HR@1 for SECURITIES from 20.51% to 61.54

The greedy weight matrix optimization approach, which creates product-specific weight configurations while optimizing for overall HR@3 performance, provided a more robust solution. This approach delivered significant improvements in HR@3 across multiple products, with particularly notable gains for organizational customers in CASH_MANAGEMENT (+37.66 percentage points), TRADE_SERVICES (+44.45 pp), and BUILDUP_F (+14.28 pp). Even more remarkably, despite optimizing for HR@3, the approach also delivered substantial improvements in HR@1 for several critical products, including INVESTMENT_FUNDS (+60.72 pp for organizational customers) and SECURITIES (+38.46 pp for individual customers).

The counterintuitive finding that some low-prevalence products achieved optimal performance with full weight on a single approach—often not the propensity model, despite its strong individual performance—highlights the complex relationship between product characteristics, class distribution, and optimal recommendation strategy. For example, FOREX, SECURITIES, and CASH_MANAGEMENT for individual customers received 100% weight on Customer Similarity, while BUILDUP_F and TRADE_SERVICES for organizational customers similarly relied entirely on Customer Similarity. This suggests that while propensity models provide the best general-purpose solution, a more nuanced approach is beneficial for specific product categories with unique characteristics.

The matrix-based weight optimization revealed that balanced weights (approximately 0.33 for each model) worked best for common products like BB and DEPOSITS, while specialized products benefited from more focused strategies. This pattern was consistent across both individual and organizational customer segments, suggesting fundamental differences in recommendation dynamics based on product characteristics rather than customer type.

The ability to optimize weights at the product level provides flexibility that could become increasingly valuable as more diverse data sources become available. Our framework can easily incorporate additional recommendation signals—such as customer click behavior, website engagement, or external economic indicators—by extending the weight matrix to include these new dimensions. This extensibility represents a significant advantage over simpler ensemble approaches and positions our methodology for continued relevance as recommendation systems grow more sophisticated.

In conclusion, our research demonstrates that while machine learning-based propensity models generally outperform similarity-based approaches for banking product recommendations, significant complementary information exists across different recommendation strategies. Our matrix-based ensemble methodology successfully leverages these complementary strengths through certain product-specific weight configurations, delivering substantial performance improvements for multiple product categories across both individual and organizational customers. By optimizing for HR@3 while maintaining or improving HR@1 for many products, our approach balances precision and recall considerations effectively, ensuring that customers receive relevant recommendations that appear within the top positions. This balance is crucial for practical implementation, where recommendation interfaces may display only a limited number of suggestions.

CHAPTER VI

Limitations and Future Work

This chapter examines the constraints of the current research approach and outlines promising avenues for advancing banking recommendation systems. Rather than treating limitations and future work as separate concerns, we present them together to highlight how future research can directly address the challenges we encountered.

6.1 Dataset Limitations and Enhanced Data Resources

The provided dataset was relatively small, particularly for certain product categories. Some specialized products had fewer than 10 examples in the test set, making performance metrics for these categories less reliable. Additionally, the dataset may not fully represent the broader customer population, potentially limiting generalizability across diverse banking environments.

Future research should focus on acquiring larger, more representative datasets spanning longer time periods, particularly for low-prevalence products. Expanding the feature set to include detailed customer attributes and banking behaviors would provide richer signals for all recommendation models. Larger datasets would not only improve model training but also enable more reliable performance assessment for specialized banking products.

Future work should explicitly model the temporal progression of customer product adoption. This would include sequence-aware recommendation techniques that consider the order and timing of previous purchases. Time-toevent models could predict not just what product a customer might adopt next, but also when they are likely to be receptive to recommendations, creating opportunities for more timely and relevant offers.

6.2 Validation Methodology

The validation approach randomly withheld products rather than specifically hiding the most recently acquired product. A more naturalistic validation would simulate real-world recommendation scenarios by testing the model's ability to predict a customer's next product acquisition based on their historical progression. This limitation, combined with the train-test split rather than cross-validation, creates uncertainty about how our system would perform in genuine forward-looking recommendation scenarios.

Implementing temporally-aware validation protocols would provide more realistic performance assessment. This includes testing models on predicting customers' most recent product acquisitions based on their prior history, as well as employing rolling-window validation to simulate how models would perform over time. Additionally, k-fold cross-validation with nested validation for weight optimization would yield more reliable and unbiased performance estimates.

6.3 Weight Optimization and Advanced Techniques

The weights generated through the greedy search algorithm were effectively overfitted to the testing set, the best-performing weights were selected on the same data used for evaluation. This likely overestimates any performance advantage of the ensemble system compared to what would be observed with independent validation data. The lack of a separate hold-out set for weight optimization represents a methodological weakness.

Exploring more sophisticated approaches to weight optimization, such as Bayesian optimization or reinforcement learning, could yield better weight configurations while reducing the risk of overfitting. Multi-objective optimization could explicitly balance performance across different metrics and product categories, aligning recommendations more closely with business objectives. These advanced techniques would provide more robust weight configurations that generalize better to new customers.

6.4 Contextual Information and Additional Signals

Beyond temporal patterns, the models did not incorporate other potentially valuable information such as transaction history, product usage intensity, or economic context. The absence of this information likely limits the models' ability to capture nuanced aspects of customer behavior and needs.

Leveraging the ensemble framework to incorporate diverse data sources would enhance recommendation relevance. This includes customer digital engagement (website activity, mobile app usage), cross-channel interactions, and external economic indicators. Weight optimization approach provides a flexible foundation for integrating these diverse signals in a manner customized to each product category.

6.5 Class Imbalance

The severe class imbalance in our dataset, particularly for specialized products, represents a fundamental challenge that affected both model training and evaluation. While the propensity model showed some resilience to this issue, all models struggled with low-prevalence products. This imbalance reflects real-world product adoption patterns but complicates meaningful performance assessment, particularly for niche but potentially high-value products.

6.6 Real-world Validation

While testing in a controlled environment provides valuable insights, it cannot fully capture how these recommendations would perform in actual banking environments. Customer receptivity to recommendations, implementation challenges, and business impact remain theoretical until tested in practice.

Conducting controlled trials in actual banking environments would provide invaluable insights into real-world effectiveness. A/B testing different recommendation strategies could validate the performance advantages observed in offline evaluation, while providing insights into practical implementation challenges and customer response patterns.

6.7 Conclusion

The limitations identified in this research highlight important opportunities for advancing banking recommendation systems. While the findings of this study suggest that propensity models currently provide the strongest baseline performance, the matrix-based ensemble methodology we've developed offers a valuable framework for future extensions. As banking continues to evolve toward more personalized customer experiences, recommendation systems that address the challenges and opportunities outlined in this chapter will play an increasingly critical role in matching customers with financial products that best meet their needs.

CHAPTER VII

Appendix

Product Category	Class 0 (Individual)	Class 1 (Individual)	Class 0 (Org)	Class 1 (Org)
BB	20 (0.47%)	4,226 (99.53%)	111 (0.98%)	11,189 (99.02%)
DEPOSITS	117 (2.76%)	4,129 (97.24%)	164 (1.45%)	11,136 (98.55%)
RETAIL_PRODUCTS	1,143 (26.91%)	3,103 (73.09%)	11,105 (98.27%)	195 (1.73%)
CORPORATE_LOANS	2,116 (49.82%)	2,130 (50.18%)	10,978 (97.16%)	322 (2.84%)
REMITTANCE	2,409 (56.74%)	1,837 (43.26%)	9,118 (80.68%)	2,182 (19.32%)
INSURANCE	2,613 (61.54%)	1,633 (38.46%)	11,156 (98.73%)	144 (1.27%)
BUILDUP_F	3,366 (79.27%)	880 (20.73%)	11,141 (98.59%)	159 (1.41%)
INVESTMENTS_AND_SECURITIES	3,613 (85.10%)	633 (14.90%)	11,147 (98.64%)	153 (1.36%)
INVESTMENT_FUNDS	3,668 (86.40%)	578 (13.60%)	11,165 (98.81%)	135 (1.19%)
FOREX	4,021 (94.71%)	225 (5.29%)	11,146 (98.64%)	154 (1.36%)
SECURITIES	4,094 (96.41%)	152 (3.59%)	11,272 (99.75%)	28 (0.25%)
CASH_MANAGEMENT	4,210 (99.15%)	36 (0.85%)	10,456 (92.55%)	844 (7.45%)
RETAIL_LOAN	4,232 (99.67%)	14 (0.33%)	11,298 (99.98%)	2 (0.02%)
TRADE_SERVICES	4,233 (99.69%)	13 (0.31%)	11,254 (99.59%)	46 (0.41%)
CORPORATE_CARDS	4,245 (99.98%)	1 (0.02%)	11,244 (99.50%)	56 (0.50%)
CORPORATE_FINANCE	4,246 (100%)	0 (0.00%)	11,300 (100%)	0 (0.00%)

Table A.1: Class Distribution of Product Ownership by Category (Individual and Org)

Product Category	Class 0 (Non-owners)	Class 1 (Owners)
BB	20	4,226
DEPOSITS	117	4,129
RETAIL_PRODUCTS	1,143	3,103
CORPORATE_LOANS	2,116	2,130
REMITTANCE	2,409	1,837
INSURANCE	2,613	1,633
BUILDUP_F	3,366	880
INVESTMENTS_AND_SECURITIES	3,613	633
INVESTMENT_FUNDS	3,668	578
FOREX	4,021	225
SECURITIES	4,094	152
CASH_MANAGEMENT	4,210	36
RETAIL_LOAN	4,232	14
TRADE_SERVICES	4,233	13
CORPORATE_CARDS	4,245	1
CORPORATE_FINANCE	4,246	0

Table A.2: Class Distribution of Product Ownership by Category (Individual)

Product Category	**Class 0 (Non-owners)**	**Class 1 (Owners)**
BB	111	11,189
DEPOSITS	164	11,136
REMITTANCE	9,118	2,182
CASH_MANAGEMENT	10,456	844
CORPORATE_LOANS	10,978	322
RETAIL_PRODUCTS	11,105	195
BUILDUP_F	11,141	159
FOREX	11,146	154
INVESTMENTS_AND_SECURITIES	11,147	153
INSURANCE	11,156	144
INVESTMENT_FUNDS	11,165	135
CORPORATE_CARDS	11,244	56
TRADE_SERVICES	11,254	46
SECURITIES	11,272	28
RETAIL_LOAN	11,298	2
CORPORATE_FINANCE	11,300	0

Table A.3: Class Distribution of Product Ownership by Category (Organization)

Product Category	Class 0 (Combined)	Class 1 (Combined)
BB	131 (0.84%)	15,415 (99.16%)
DEPOSITS	281 (1.81%)	15,265 (98.19%)
RETAIL_PRODUCTS	12,248 (78.83%)	3,298 (21.17%)
CORPORATE_LOANS	13,094 (84.27%)	2,452 (15.73%)
REMITTANCE	11,527 (74.15%)	4,019 (25.85%)
INSURANCE	13,769 (88.61%)	1,777 (11.39%)
BUILDUP_F	14,507 (93.44%)	1,039 (6.56%)
INVESTMENTS_AND_SECURITIES	14,760 (95.08%)	786 (4.92%)
INVESTMENT_FUNDS	14,833 (95.71%)	713 (4.29%)
FOREX	15,167 (97.62%)	379 (2.38%)
SECURITIES	15,366 (98.64%)	180 (1.16%)
CASH_MANAGEMENT	14,666 (94.13%)	880 (5.87%)
RETAIL_LOAN	15,530 (99.74%)	16 (0.26%)
TRADE_SERVICES	15,487 (99.65%)	59 (0.35%)
CORPORATE_CARDS	15,489 (99.74%)	57 (0.36%)
CORPORATE_FINANCE	15,546 (100%)	0 (0.00%)

Table A.4: Class Distribution of Product Ownership by Category (Combined)

Income Source	Count
BUSINESS	2,745
SALARY	1,058
PENSION	205
NO_DATA	54
ALLOWANCE	53
REMITTANCE	43
COMMISSION	34
OTHER_SOURCES_NOT_SPECIFIED	28
INTEREST_SAVINGS_PLACEMENTS_INVESTMENTS	25
ECONOMICALLY_INACTIVE	1

Table A.5: Value Counts for Income Source (Individual)

Business Owner	Count
Y	3,241
N	1,005

Table A.6: Value Counts for Business Owner (Individual)

Region	Count
NATIONAL CAPITAL REGION	1,456
REGION IV-A (CALABARZON)	647
REGION III (CENTRAL LUZON)	423
REGION VII (WESTERN VISAYAS)	323
REGION VI (CENTRAL VISAYAS)	209
...	...

Table A.7: Value Counts for Region (Individual)

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
BB	0.991968	0.882642	0.910612	0.198394	0.795181	0.958501	2241
DEPOSITS	0.996861	0.886510	0.914688	0.199372	0.800448	0.962780	2230
RETAIL_PRODUCTS	0.375000	0.375000	0.375000	0.075000	0.375000	0.375000	40
REMITTANCE	1.000000	0.967111	0.975494	0.200000	0.941889	0.987893	413
INSURANCE	0.200000	0.056000	0.090948	0.040000	0.000000	0.120000	25
BUILDUP_F	0.942857	0.354286	0.501016	0.188571	0.057143	0.885714	35
CORPORATE_CARDS	0.750000	0.244444	0.368535	0.150000	0.000000	0.500000	12
FOREX	0.966667	0.821111	0.855945	0.193333	0.766667	0.833333	30
CASH_MANAGEMENT	1.000000	0.435288	0.579768	0.200000	0.006173	0.975309	162
CORPORATE_LOANS	1.000000	0.257576	0.436979	0.200000	0.000000	0.090909	66
INVESTMENTS_AND_SECURITIES	0.928571	0.803571	0.836304	0.185714	0.678571	0.928571	28
INVESTMENT_FUNDS	0.928571	0.595238	0.681113	0.185714	0.321429	0.928571	28
TRADE_SERVICES	0.777778	0.481481	0.558191	0.155556	0.222222	0.777778	9
Overall	0.984584	0.853867	0.887102	0.196916	0.757097	0.941154	5319

Table A.8: Performance metrics (Hit Rate, MRR, NDCG, Precision, HR@1, and HR@3) for product recommendations using the Customer Similarity Only approach for organizational customers

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
BB	1.0000	0.9931	0.9949	0.2000	0.9866	1.0000	2241
DEPOSITS	1.0000	1.0000	1.0000	0.2000	1.0000	1.0000	2230
RETAIL_PRODUCTS	0.8500	0.2767	0.4190	0.1700	0.0000	0.7000	40
REMITTANCE	1.0000	0.9988	0.9991	0.2000	0.9976	1.0000	413
INSURANCE	0.7600	0.2287	0.3574	0.1520	0.0000	0.3600	25
BUILDUP_F	0.8000	0.3081	0.4320	0.1600	0.0000	0.6857	35
CORPORATE_CARDS	0.5833	0.2153	0.3077	0.1167	0.0000	0.5000	12
FOREX	0.8667	0.4261	0.5391	0.1733	0.0667	0.8000	30
CASH_MANAGEMENT	0.9383	0.5239	0.6237	0.1877	0.3827	0.4074	162
CORPORATE_LOANS	1.0000	0.4033	0.5537	0.2000	0.0000	0.6970	66
INVESTMENTS_AND_SECURITIES	1.0000	0.8214	0.8682	0.2000	0.6429	1.0000	28
INVESTMENT_FUNDS	0.9286	0.7679	0.8099	0.1857	0.6071	0.9286	28
TRADE_SERVICES	0.8889	0.3500	0.4808	0.1778	0.1111	0.3333	9
Overall	0.9923	0.9532	0.9629	0.1985	0.9312	0.9671	5319

Table A.9: Performance metrics (Hit Rate, MRR, NDCG, Precision, HR@1, and HR@3) for product recommendations using the Product Similarity Only approach for organizational customers.

Product	**Hit Rate**	**MRR**	**NDCG**	**Precision**	**HR@1**	**HR@3**	**Total Tested**
BB	1.0000	0.9993	0.9995	0.2000	0.9987	1.0000	2241
DEPOSITS	1.0000	0.9991	0.9993	0.2000	0.9982	1.0000	2230
RETAIL_PRODUCTS	0.8250	0.5925	0.6514	0.1650	0.4250	0.7500	40
REMITTANCE	1.0000	0.8909	0.9173	0.2000	0.8499	0.8717	413
INSURANCE	0.7600	0.3467	0.4475	0.1520	0.1600	0.4800	25
BUILDUP_F	0.8286	0.5281	0.6041	0.1657	0.3143	0.7429	35
CORPORATE_CARDS	0.5833	0.3056	0.3770	0.1167	0.0833	0.5833	12
FOREX	0.8667	0.7139	0.7518	0.1733	0.6333	0.7667	30
CASH_MANAGEMENT	0.9259	0.3137	0.4648	0.1852	0.0370	0.6111	162
CORPORATE_LOANS	1.0000	0.6869	0.7653	0.2000	0.5000	0.8485	66
INVESTMENTS_AND_SECURITIES	1.0000	1.0000	1.0000	0.2000	1.0000	1.0000	28
INVESTMENT_FUNDS	0.9286	0.6250	0.7045	0.1857	0.3214	0.9286	28
TRADE_SERVICES	0.8889	0.3500	0.4808	0.1778	0.1111	0.3333	9
Overall	0.9919	0.9506	0.9608	0.1984	0.9295	0.9665	5319

Table A.10: Performance metrics (Hit Rate, MRR, NDCG, Precision, HR@1, and HR@3) for product recommendations using the Propensity Model Only approach for organizational customers.

Product	Customer Similarity	Product Similarity	Propensity Model	Total Tested
BB	0.991968	1.0000	1.0000	2241
DEPOSITS	0.996861	1.0000	1.0000	2230
RETAIL_PRODUCTS	0.375000	0.8500	0.8250	40
REMITTANCE	1.000000	1.0000	1.0000	413
INSURANCE	0.200000	0.7600	0.7600	25
BUILDUP_F	0.942857	0.8000	0.8286	35
CORPORATE_CARDS	0.750000	0.5833	0.5833	12
FOREX	0.966667	0.8667	0.8667	30
CASH_MANAGEMENT	1.000000	0.9383	0.9259	162
CORPORATE_LOANS	1.000000	1.0000	1.0000	66
INVESTMENTS_AND_SECURITIES	0.928571	1.0000	1.0000	28
INVESTMENT_FUNDS	0.928571	0.9286	0.9286	28
TRADE_SERVICES	0.777778	0.8889	0.8889	9
Average	**0.847354**	**0.9233**	**0.9236**	**-**

Table A.11: Hit rate comparison across Customer Similarity, Product Similarity, and Propensity Model approaches for organizational customers.

Product	Customer Similarity	Product Similarity	Propensity Model	Total Tested
BB	0.882642	0.9931	0.9993	2241
DEPOSITS	0.886510	1.0000	0.9991	2230
RETAIL_PRODUCTS	0.375000	0.2767	0.5925	40
REMITTANCE	0.967111	0.9988	0.8909	413
INSURANCE	0.056000	0.2287	0.3467	25
BUILDUP_F	0.354286	0.3081	0.5281	35
CORPORATE_CARDS	0.244444	0.2153	0.3056	12
FOREX	0.821111	0.4261	0.7139	30
CASH_MANAGEMENT	0.435288	0.5239	0.3137	162
CORPORATE_LOANS	0.257576	0.4033	0.6869	66
INVESTMENTS_AND_SECURITIES	0.803571	0.8214	1.0000	28
INVESTMENT_FUNDS	0.595238	0.7679	0.6250	28
TRADE_SERVICES	0.481481	0.3500	0.3500	9
Average	**0.5520**	**0.6111**	**0.6882**	**-**

Table A.12: MRR comparison across Customer Similarity, Product Similarity, and Propensity Model approaches for organizational customers.

Product	Customer Similarity	Product Similarity	Propensity Model	Total Tested
BB	0.795181	0.9866	0.9987	2241
DEPOSITS	0.800448	1.0000	0.9982	2230
RETAIL_PRODUCTS	0.375000	0.0000	0.4250	40
REMITTANCE	0.941889	0.9976	0.8499	413
INSURANCE	0.000000	0.0000	0.1600	25
BUILDUP_F	0.057143	0.0000	0.3143	35
CORPORATE_CARDS	0.000000	0.0000	0.0833	12
FOREX	0.766667	0.0667	0.6333	30
CASH_MANAGEMENT	0.006173	0.3827	0.0370	162
CORPORATE_LOANS	0.000000	0.0000	0.5000	66
INVESTMENTS_AND_SECURITIES	0.678571	0.6429	1.0000	28
INVESTMENT_FUNDS	0.321429	0.6071	0.3214	28
TRADE_SERVICES	0.222222	0.1111	0.1111	9
Average	**0.430**	**0.475**	**0.509**	**-**

Table A.13: Hit Rate @1 comparison across Customer Similarity, Product Similarity, and Propensity Model approaches for organizational customers.

Product	Customer Similarity	Product Similarity	Propensity Model	Total Tested
BB	0.958501	1.0000	1.0000	2241
DEPOSITS	0.962780	1.0000	1.0000	2230
RETAIL_PRODUCTS	0.375000	0.7000	0.7500	40
REMITTANCE	0.987893	1.0000	0.8717	413
INSURANCE	0.120000	0.3600	0.4800	25
BUILDUP_F	0.885714	0.6857	0.7429	35
CORPORATE_CARDS	0.500000	0.5000	0.5833	12
FOREX	0.833333	0.8000	0.7667	30
CASH_MANAGEMENT	0.975309	0.4074	0.6111	162
CORPORATE_LOANS	0.090909	0.6970	0.8485	66
INVESTMENTS_AND_SECURITIES	0.928571	1.0000	1.0000	28
INVESTMENT_FUNDS	0.928571	0.9286	0.9286	28
TRADE_SERVICES	0.777778	0.3333	0.3333	9
Average	**0.8472**	**0.7790**	**0.8203**	**-**

Table A.14: HR@3 comparison across Customer Similarity, Product Similarity, and Propensity Model approaches for organizational customers.

Product Category	Customer Similarity	Product Similarity	Propensity
BB	0.30	0.50	0.20
DEPOSITS	0.10	0.90	0.00
RETAIL_PRODUCTS	0.00	0.20	0.80
REMITTANCE	0.10	0.50	0.40
INSURANCE	0.10	0.20	0.70
BUILDUP_F	0.30	0.00	0.70
CORPORATE_CARDS	0.60	0.00	0.40
FOREX	1.00	0.00	0.00
CASH_MANAGEMENT	0.10	0.90	0.00
CORPORATE_LOANS	0.00	0.00	1.00
INVESTMENTS_AND_SECURITIES	0.10	0.70	0.20
INVESTMENT_FUNDS	0.70	0.30	0.00
TRADE_SERVICES	1.00	0.00	0.00

Table A.15: Per-Product Optimized Model Weights for Hitrate@1 (Organizational)

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3
BB	0.9946	0.9906	0.9917	0.1989	0.9866	0.9946
DEPOSITS	1.0000	1.0000	1.0000	0.2000	1.0000	1.0000
RETAIL_PRODUCTS	0.4000	0.1333	0.2000	0.0800	0.0000	0.4000
REMITTANCE	1.0000	0.9988	0.9991	0.2000	0.9976	1.0000
INSURANCE	0.2400	0.0680	0.1096	0.0480	0.0000	0.0400
BUILDUP_F	0.9429	0.3090	0.4667	0.1886	0.0000	0.8571
CORPORATE_CARDS	0.7500	0.2306	0.3570	0.1500	0.0000	0.3333
FOREX	0.9667	0.4822	0.6055	0.1933	0.1000	0.8333
CASH_MANAGEMENT	1.0000	0.6851	0.7671	0.2000	0.3827	0.9815
CORPORATE_LOANS	0.9848	0.2508	0.4278	0.1970	0.0000	0.0909
INVESTMENTS_AND_SECURITIES	1.0000	0.8214	0.8682	0.2000	0.6429	1.0000
INVESTMENT_FUNDS	0.9286	0.7679	0.8099	0.1857	0.6071	0.9286
TRADE_SERVICES	0.7778	0.3556	0.4610	0.1556	0.1111	0.6667

Table A.16: Performance metrics per product using weights optimized for Hitrate@1 of Cash Management (Organizational).

Product	Propensity	Product Similarity	Customer Similarity
BB	0.33	0.33	0.33
BUILDUP_F	0.00	0.00	1.00
CASH_MANAGEMENT	0.00	0.40	0.60
CORPORATE_CARDS	0.20	0.50	0.30
CORPORATE_LOANS	0.70	0.00	0.30
DEPOSITS	0.33	0.33	0.33
FOREX	0.20	0.70	0.10
INSURANCE	1.00	0.00	0.00
INVESTMENTS_AND_SECURITIES	0.33	0.33	0.33
INVESTMENT_FUNDS	0.33	0.33	0.33
REMITTANCE	0.33	0.33	0.33
RETAIL_PRODUCTS	1.00	0.00	0.00
TRADE_SERVICES	0.00	0.00	1.00

Table A.17: Optimized Model Weights from Greedy Search for Organizational Customers (HitRate@3)

Product	Hit Rate	MRR	NDCG	HR@1	HR@3
BB	1.000000	0.999331	0.999506	0.998661	1.000000
DEPOSITS	1.000000	0.999327	0.999503	0.998655	1.000000
REMITTANCE	1.000000	0.996368	0.997319	0.992736	1.000000
CASH_MANAGEMENT	1.000000	0.652675	0.743151	0.320988	0.987654
CORPORATE_LOANS	1.000000	0.638131	0.729922	0.378788	0.878788
INVESTMENTS_AND_SECURITIES	1.000000	0.982143	0.986819	0.964286	1.000000
BUILDUP_F	0.942857	0.354286	0.501016	0.057143	0.885714
INVESTMENT_FUNDS	0.928571	0.928571	0.928571	0.928571	0.928571
FOREX	0.866667	0.450000	0.558317	0.066667	0.866667
CORPORATE_CARDS	0.833333	0.304167	0.435675	0.000000	0.583333
RETAIL_PRODUCTS	0.825000	0.592500	0.651438	0.425000	0.750000
TRADE_SERVICES	0.777778	0.481481	0.558191	0.222222	0.777778
INSURANCE	0.760000	0.346667	0.447536	0.160000	0.480000

Table A.18: Ensemble System Performance for Organizational Customers

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
INSURANCE	1.000000	0.670603	0.755168	0.200000	0.404908	0.957055	326
BB	1.000000	0.997642	0.998259	0.200000	0.995283	1.000000	848
DEPOSITS	1.000000	0.999400	0.999557	0.200000	0.998801	1.000000	834
RETAIL_PRODUCTS	1.000000	0.989533	0.992274	0.200000	0.979066	1.000000	621
INVESTMENTS_AND_SECURITIES	1.000000	0.500864	0.627019	0.200000	0.170370	0.866667	135
REMITTANCE	1.000000	0.753991	0.817564	0.200000	0.543662	1.000000	355
INVESTMENT_FUNDS	1.000000	0.474727	0.607811	0.200000	0.122951	0.909836	122
CORPORATE_LOANS	1.000000	0.853970	0.892042	0.200000	0.714623	0.997642	424
BUILDUP_F	1.000000	0.490952	0.618412	0.200000	0.165000	0.874286	175
FOREX	0.627451	0.200327	0.304165	0.125490	0.019608	0.274510	51
SECURITIES	0.846154	0.342735	0.466232	0.169231	0.102564	0.487179	39
CASH_MANAGEMENT	0.444444	0.088889	0.171935	0.088889	0.000000	0.000000	9
Overall	0.992384	0.857629	0.891963	0.198477	0.757553	0.966489	3939

Table A.19: Performance metrics (Hit Rate, MRR, NDCG, Precision, HR@1, and HR@3) for product recommendations using the Product Similarity Only approach for individual customers.

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
INSURANCE	0.993865	0.516207	0.634382	0.198773	0.269939	0.720859	326
BB	0.839623	0.532115	0.609251	0.167925	0.336085	0.705189	848
DEPOSITS	0.875300	0.643465	0.701772	0.175060	0.485010	0.788969	834
RETAIL_PRODUCTS	0.942029	0.520719	0.625181	0.188406	0.294686	0.724499	621
INVESTMENTS_AND_SECURITIES	0.925926	0.869383	0.883904	0.185185	0.822222	0.918519	135
REMITTANCE	0.971831	0.573756	0.672614	0.194366	0.338028	0.684507	355
INVESTMENT_FUNDS	0.983607	0.954918	0.962430	0.196721	0.926230	0.983607	122
CORPORATE_LOANS	0.917453	0.580267	0.666436	0.183491	0.308962	0.801887	424
BUILDUP_F	0.891429	0.497238	0.597128	0.178286	0.217143	0.782857	175
FOREX	0.862745	0.343791	0.471705	0.172549	0.078431	0.607843	51
SECURITIES	1.000000	0.773504	0.831801	0.200000	0.588744	1.000000	39
CASH_MANAGEMENT	0.666667	0.222222	0.332476	0.133333	0.000000	0.444444	9
Overall	0.907591	0.583862	0.6649012	0.174392	0.381569	0.757044	3939

Table A.20: Performance metrics (Hit Rate, MRR, NDCG, Precision, HR@1, and HR@3) for product recommendations using the Customer Only approach for individual customers.

Product	Hit Rate	MRR	NDCG	Precision	HR@1	HR@3	Total Tested
INSURANCE	1.000000	0.810583	0.859288	0.200000	0.656442	0.984663	326
BB	1.000000	1.000000	1.000000	0.200000	1.000000	1.000000	848
DEPOSITS	1.000000	0.998801	0.999115	0.200000	0.997602	1.000000	834
RETAIL_PRODUCTS	1.000000	0.959742	0.970246	0.200000	0.921095	1.000000	621
INVESTMENTS_AND_SECURITIES	1.000000	0.996296	0.997266	0.200000	0.992593	1.000000	135
REMITTANCE	1.000000	0.868873	0.902768	0.200000	0.754930	0.994366	355
INVESTMENT_FUNDS	0.983607	0.934426	0.947110	0.196721	0.893443	0.983607	122
CORPORATE_LOANS	1.000000	0.875904	0.907740	0.200000	0.773585	0.974057	424
BUILDUP_F	0.994286	0.667810	0.750963	0.198857	0.428571	0.931429	175
FOREX	0.960784	0.437582	0.565399	0.192157	0.196078	0.588235	51
SECURITIES	1.000000	0.544444	0.660350	0.200000	0.205128	0.897436	33
CASH_MANAGEMENT	1.000000	0.344444	0.502435	0.200000	0.111111	0.444444	9
Overall	0.998731	0.922340	0.941859	0.184392	0.862909	0.984259	3939

Table A.21: Performance metrics (Hit Rate, MRR, NDCG, Precision, HR@1, and HR@3) for product recommendations using the Propensity Model Only approach for individual customers.

Product Category	Customer Similarity	Product Similarity	Propensity
INSURANCE	0.00	0.20	0.80
BB	0.10	0.70	0.20
DEPOSITS	0.30	0.60	0.10
RETAIL_PRODUCTS	0.10	0.60	0.30
INVESTMENTS_AND_SECURITIES	0.20	0.00	0.80
REMITTANCE	0.00	0.20	0.80
INVESTMENT_FUNDS	0.90	0.00	0.10
CORPORATE_LOANS	0.40	0.00	0.60
BUILDUP_F	0.60	0.00	0.40
FOREX	0.30	0.00	0.70
SECURITIES	0.80	0.10	0.10
CASH_MANAGEMENT	0.00	0.00	1.00

Table A.22: Per-Product Optimized Model Weights for Hitrate@1 (Individuals)

Product	Hit Rate	MRR	NCDG	Precision	HR@1	HR@3
INSURANCE	0.996933	0.673006	0.755672	0.199387	0.429448	0.957055
BB	1.000000	0.921384	0.941672	0.200000	0.854953	0.997642
DEPOSITS	1.000000	0.913869	0.936224	0.200000	0.835731	0.997602
RETAIL_PRODUCTS	1.000000	0.887815	0.916874	0.200000	0.789050	1.000000
INVESTMENTS_AND_SECURITIES	1.000000	0.931975	0.948811	0.200000	0.896296	0.962963
REMITTANCE	1.000000	0.688920	0.767513	0.200000	0.478873	0.901408
INVESTMENT_FUNDS	0.991803	0.985246	0.986777	0.198361	0.983607	0.983607
CORPORATE_LOANS	0.992925	0.835810	0.875108	0.198585	0.738208	0.919811
BUILDUP_F	0.960000	0.532476	0.640784	0.192000	0.222857	0.794286
FOREX	0.725490	0.319608	0.420407	0.145098	0.098039	0.529412
SECURITIES	1.000000	0.788462	0.842844	0.200000	0.615385	0.974359
CASH_MANAGEMENT	0.666667	0.198148	0.312201	0.133333	0.000000	0.222222
Overall	0.992638	0.838085	0.877462	0.160934	0.7220107	0.958872

Table A.23: Performance metrics per product using weights optimized for Hitrate@1 of Securities (Individuals).

Product Category	Customer Similarity	Product Similarity	Propensity
INSURANCE	0.00	0.40	0.60
BB	0.60	0.40	0.00
DEPOSITS	0.00	1.00	0.00
RETAIL_PRODUCTS	0.80	0.20	0.00
INVESTMENTS_AND_SECURITIES	0.70	0.00	0.30
REMITTANCE	0.50	0.50	0.00
INVESTMENT_FUNDS	1.00	0.00	0.00
CORPORATE_LOANS	0.30	0.50	0.20
BUILDUP_F	0.00	0.00	1.00
FOREX	1.00	0.00	0.00
SECURITIES	1.00	0.00	0.00
CASH_MANAGEMENT	1.00	0.00	0.00

Table A.24: Per-Product Optimized Model Weights for Hitrate@3 (Individuals)

Product	Customer Similarity	Product Similarity	Propensity Model
INSURANCE	0.2000	0.1000	0.7000
BB	0.3333	0.3333	0.3333
DEPOSITS	0.3333	0.3333	0.3333
RETAIL_PRODUCTS	0.3333	0.3333	0.3333
INVESTMENTS_AND_SECURITIES	0.3333	0.3333	0.3333
REMITTANCE	0.5000	0.5000	0.0000
INVESTMENT_FUNDS	0.3333	0.3333	0.3333
CORPORATE_LOANS	0.3000	0.5000	0.2000
BUILDUP_F	0.0000	0.0000	1.0000
FOREX	1.0000	0.0000	0.0000
SECURITIES	1.0000	0.0000	0.0000
CASH_MANAGEMENT	1.0000	0.0000	0.0000

Table A.25: Optimized Model Weights from Greedy Search for Individual Customers

Product	Hit Rate	MRR	NDCG	HR@1	HR@3
INSURANCE	1.000000	0.811500	0.859962	0.656442	0.987730
BB	1.000000	0.998821	0.999130	0.997642	1.000000
DEPOSITS	1.000000	0.998801	0.999115	0.997602	1.000000
RETAIL_PRODUCTS	1.000000	0.991143	0.994463	0.982287	1.000000
INVESTMENTS_AND_SECURITIES	1.000000	0.969136	0.977159	0.940741	1.000000
REMITTANCE	1.000000	0.711668	0.786596	0.467606	1.000000
INVESTMENT_FUNDS	1.000000	0.963115	0.972516	0.934426	0.983607
CORPORATE_LOANS	1.000000	0.842374	0.883501	0.691038	1.000000
BUILDUP_F	0.994286	0.667610	0.770963	0.428571	0.931429
FOREX	0.862745	0.343791	0.477765	0.078431	0.607843
SECURITIES	1.000000	0.773504	0.831801	0.589744	1.000000
CASH_MANAGEMENT	0.666667	0.222222	0.332476	0.000000	0.444444

Table A.26: Ensemble System Performance for Individual Customers

Product	Customer Similarity		Product Similarity		Propensity Model		Total Tested
	HR@1	HR@3	HR@1	HR@3	HR@1	HR@3	
INSURANCE	0.269939	0.720859	0.404908	0.957055	0.656442	0.984663	326
BB	0.336085	0.705189	0.995283	1.000000	1.000000	1.000000	848
DEPOSITS	0.485010	0.788969	0.998801	1.000000	0.997602	1.000000	834
RETAIL_PRODUCTS	0.294686	0.724499	0.979066	1.000000	0.921095	1.000000	621
INVESTMENTS_AND_SECURITIES	0.822222	0.918519	0.170370	0.866667	0.992593	1.000000	135
REMITTANCE	0.338028	0.684507	0.543662	1.000000	0.754930	0.994366	355
INVESTMENT_FUNDS	0.926230	0.983607	0.122951	0.909836	0.893443	0.983607	122
CORPORATE_LOANS	0.308962	0.801887	0.714623	0.997642	0.773585	0.974057	424
BUILDUP_F	0.217143	0.782857	0.165000	0.874286	0.428571	0.931429	175
FOREX	0.078431	0.607843	0.019608	0.274510	0.196078	0.588235	51
SECURITIES	0.588744	1.000000	0.102564	0.487179	0.205128	0.897436	39
CASH_MANAGEMENT	0.000000	0.444444	0.000000	0.000000	0.111111	0.444444	9
Overall	**0.381569**	**0.757044**	**0.757553**	**0.966489**	**0.862909**	**0.984259**	**3939**

Table A.27: Comparison of HR@1 and HR@3 across all models for individual customers.

Product	Customer Similarity		Product Similarity		Propensity Model		Total Tested
	HR@1	HR@3	HR@1	HR@3	HR@1	HR@3	
BB	0.795181	0.958501	0.9866	1.0000	0.9987	1.0000	2241
DEPOSITS	0.800448	0.962780	1.0000	1.0000	0.9982	1.0000	2230
RETAIL_PRODUCTS	0.375000	0.375000	0.0000	0.7000	0.4250	0.7500	40
REMITTANCE	0.941889	0.987893	0.9976	1.0000	0.8499	0.8717	413
INSURANCE	0.000000	0.120000	0.0000	0.3600	0.1600	0.4800	25
BUILDUP_F	0.057143	0.885714	0.0000	0.6857	0.3143	0.7429	35
CORPORATE_CARDS	0.000000	0.500000	0.0000	0.5000	0.0833	0.5833	12
FOREX	0.766667	0.833333	0.0667	0.8000	0.6333	0.7667	30
CASH_MANAGEMENT	0.006173	0.975309	0.3827	0.4074	0.0370	0.6111	162
CORPORATE_LOANS	0.000000	0.090909	0.0000	0.6970	0.5000	0.8485	66
INVESTMENTS_AND_SECURITIES	0.678571	0.928571	0.6429	1.0000	1.0000	1.0000	28
INVESTMENT_FUNDS	0.321429	0.928571	0.6071	0.9286	0.3214	0.9286	28
TRADE_SERVICES	0.222222	0.777778	0.1111	0.3333	0.1111	0.3333	9
Overall	**0.757097**	**0.941154**	**0.9312**	**0.9671**	**0.9295**	**0.9665**	**5319**

Table A.28: Comparison of HR@1 and HR@3 across all models for organizational customers.

BIBLIOGRAPHY

[1] ABBASS, Z. E. Implementing a bank sales analytics solution and a predictive model for the next best offer. Master's thesis, Nova Information Management School, 2018.

[2] ADITYA, P., BUDI, I., AND MUNAJAT, Q. A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for e-commerce in indonesia: A case study pt x. In *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (2016), IEEE, pp. 303–308.

[3] BANGKO SENTRAL NG PILIPINAS. Classifications, Powers, and Scope of Authorities of Banks, 2024. Accessed: 2024-11-15.

[4] BANGKO SENTRAL NG PILIPINAS. Eflp 2024. https://www.bsp.gov.ph/Inclusive%20Finance/EFLP/EFLP2024_1_2b.pdf, 2024. Accessed: 2024-11-13.

[5] BOUNEFFOUF, D., BOUZEGHOUB, A., AND GANÇARSKI, A. L. A contextual-bandit algorithm for mobile context-aware recommender system. In *Neural Information Processing: 19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part III 19* (2012), Springer, pp. 324–331.

[6] BOUNEFFOUF, D., AND RISH, I. A survey on practical applications of multi-armed and contextual bandits. *CoRR abs/1904.10040* (2019).

[7] CHEN, T., AND GUESTRIN, C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), 785–794.

[8] CONGRESS OF THE PHILIPPINES. The Charter of the Al-Amanah Islamic Investment Bank of the Philippines, 1990. Accessed: 2024-11-15.

[9] CONGRESS OF THE PHILIPPINES. Rural Banks Act of 1992, 1992. Accessed: 2024-11-15.

[10] CONGRESS OF THE PHILIPPINES. Republic Act No. 7906: Thrift Banks Act of 1995, 1995. Accessed: 2024-11-15.

[11] DE GANTÈS, G. On the verge of a digital banking revolution in the philippines, 2024. Accessed: 2024-11-19.

[12] DEPARTMENT, C. C. C. S. Memory-based collaborative filtering, 2007. Accessed: 2024-11-19.

[13] FABRIZI, A. Hp sps next best offer: How to re-think your marketing. Technical white paper, HP, 2018. Accessed: 2024-10-10.

[14] FERNANDEZ, L. E. M. Recommendation system for netflix. *Knowledge and Data Engineering, IEEE Transactions on* (2018).

[15] GHASEMI, M., MOOSAVI, A. H., SORKHOH, I., AGRAWAL, A., ALZHOURI, F., AND EBRAHIMI, D. An introduction to reinforcement learning: Fundamental concepts and practical applications. *arXiv preprint arXiv:2408.07712* (2024).

[16] INVESTOPEDIA. Bank. https://www.investopedia.com/terms/b/bank.asp, n.d. Accessed: 2024-11-13.

[17] LEVINAS, C. A. An analysis of memory based collaborative filtering recommender systems with improvement proposals. Master's thesis, Universitat Politècnica de Catalunya, 2014.

[18] MOHAMMEDRAZIULLAHANSARI. What is model validation and why is it important? *GeeksforGeeks* (May 2024).

[19] PHALLE, T. S., AND BHUSHAN, S. Content based filtering and collaborative filtering: A comparative study. *Journal of Advanced Zoology 45*, S4 (2024), 96–100.

[20] PUBLICIS SAPIENT. Hyper-personalization: The key to growth in banking, 2023. Accessed: 2024-11-13.

[21] QUIN, F., WEYNS, D., GALSTER, M., AND SILVA, C. C. A/b testing: A systematic literature review. *Journal of Systems and Software 211* (2024), 112011.

[22] SARWAR, B., KARYPIS, G., KONSTAN, J., AND RIEDL, J. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web* (2001), ACM, p. 285–295.

[23] SLIVKINS, A. *Introduction to Multi-Armed Bandits*. Now Foundations and Trends, 2024. arXiv preprint.

[24] UPADHYAYA, P. User-based and item-based collaborative filtering. figure 2 in: Study of mathematical model for user-based collaborative filtering and item-based collaborative filtering. https://www.researchgate.net/figure/User-based-and-Item-based-Collaborative-Filtering_fig2_366902172. Accessed: October 26, 2023.

[25] ZHOU, Q., ZHANG, X., XU, J., AND LIANG, B. Large-scale bandit approaches for recommender systems. In *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part I 24* (2017), Springer, pp. 811–821.