universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Recommender Systems in Grocery Retail – A Multi Objective Optimization Problem

verfasst von | submitted by

Christian Jan Orlowski

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 977
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Business Analytics
Betreut von | Supervisor:	Univ.-Prof. Dr. Jan Fabian Ehmke

Inhaltsverzeichnis

Zusammenfassung

Der Lebensmitteleinzelhandel arbeitet mit Nettomargen von ein bis drei Prozent, was die Rentabilität in hohem Maße abhängig von der kumulativen Effizienz operativer Entscheidungen macht, einschließlich der Frage, welche Produkte welchen Kunden empfohlen werden sollen. Standard-Empfehlungssysteme optimieren jedoch ausschließlich auf Kundenrelevanz und behandeln jedes Produkt aus Sicht des Händlers als gleichermaßen wünschenswert, wobei Marge, Bestandsdruck, Verderblichkeit, Cross-Selling-Potenzial und strategische Prioritäten ignoriert werden. Diese Arbeit befasst sich mit dem daraus resultierenden Spannungsfeld durch die Entwicklung und empirische Evaluierung eines Frameworks zur Transformation von Standard-Empfehlungssystemen im Lebensmitteleinzelhandel in geschäftsbewusste (business-aware) Systeme, die gemeinsam Kunden- und Händlerziele bedienen.

Zwei Empfehlungsmodelle, TIFU-KNN und ReCANet, werden durch vier Integrationsmethoden erweitert: multiplikatives Scoring, Weighted-Sum Mixed Integer Programming (MIP), Epsilon-Constraint-MIP und direkte Modellanpassungen. Die Ziele des Einzelhändlers werden durch einen fünfdimensionalen business value score operationalisiert, und alle Methoden werden auf dem Dunnhumby-Datensatz unter Verwendung von Recall, NDCG und aggregiertem Geschäftswert sowie einer Analyse der Auswirkungen auf Kundenebene evaluiert.

Die Ergebnisse zeigen, dass bedeutsame Steigerungen des Geschäftswerts von 8-22 % bei moderaten Genauigkeitseinbußen von 1-7 % erreichbar sind, dass die Effizienz dieses Trade-offs jedoch weniger von der Wahl der Integrationsmethode abhängt als von den strukturellen Eigenschaften des zugrundeliegenden Modells, nämlich der Diversität der Kandidatenmenge, der Score-Varianz und der Form der Score-Verteilung. TIFU-KNN übertraf ReCANet konsistent über alle Integrationsmethoden hinweg, was darauf hindeutet, dass die Qualität der Kandidatenmenge eine Obergrenze darstellt, die durch noch so viel Post-Processing nicht überwunden werden kann. Die Integration auf Repräsentationsebene erzielte den günstigsten Trade-off und lieferte eine Steigerung des Geschäftswerts von etwa 3 % bei nahezu null Genauigkeitsverlust. Die Analyse auf Kundenebene ergab zudem, dass Post-Processing und Methoden auf Modellebene systematisch unterschiedliche Nutzersegmente beeinflussen, was heterogene Behandlungsrichtlinien motiviert. Zusammengenommen definieren diese Ergebnisse geschäftsbewusste Empfehlungen neu als eine Frage danach, *wo* in der Pipeline das Geschäftssignal einfließt, und nicht nur danach, wie.

Abstract

Grocery retail operates on net margins of one to three percent, making profitability highly sensitive to the cumulative efficiency of operational decisions, including which products to recommend to which customers. Standard recommender systems, however, optimize solely for customer relevance and treat every product as equally desirable from the retailer's perspective, ignoring margin, inventory pressure, perishability, cross-selling potential, and strategic priority. This thesis addresses the resulting tension by developing and empirically evaluating a systematic framework for transforming standard grocery recommenders into business-aware systems that jointly serve customer and retailer objectives.

Two recommender models, TIFU-KNN and ReCANet, are extended through four integration methods: multiplicative scoring, weighted-sum Mixed Integer Programming, epsilon-constraint MIP, and direct model adjustments. Retailer objectives are operationalized through a five-dimensional business value score, and all methods are evaluated on the Dunnhumby dataset using Recall, NDCG, and aggregate business value, alongside a customer-level impact analysis.

The results show that meaningful business value gains of 8–22% are achievable at modest accuracy costs of 1–7%, but that the efficiency of this trade-off depends less on the choice of integration method than on structural properties of the underlying model, namely candidate-set diversity, score variance, and score-distribution shape. TIFU-KNN consistently outperformed ReCANet across all integration methods, even under provably optimal MIP formulations, indicating that candidate-set quality imposes an upper bound that no amount of post-processing can overcome. Representation-level integration achieved the most favorable trade-off, delivering approximately 3% business value gains at near-zero accuracy cost. Customer-level analysis further revealed that post-processing and model-level methods affect systematically different user segments, which motivates heterogeneous treatment policies. Taken together, these findings reframe business-aware recommendation as a question of *where* in the pipeline the business signal enters, rather than merely how.

1 Introduction

1.1 Motivation

Grocery retail is one of the largest and most competitive industries in the world, yet it operates on characteristically thin profit margins. Net margins in conventional supermarkets typically range between one and three percent (Trejo-Pech & White, 2024). As a result, the financial viability of a grocery business depends not on any single transaction but on the cumulative efficiency of millions of small decisions made every day. Among these, which products to recommend to which customers at which moment has become increasingly consequential as digital interfaces, loyalty programs, and online grocery platforms bring algorithmic recommendation directly into the shopping experience.

Recommender systems have proven to be powerful tools for personalizing these interactions. By analyzing purchase histories and identifying patterns across large customer bases, they can surface products that customers are likely to buy, reducing shopping friction and increasing basket sizes (Roy & Dutta, 2022). The academic and commercial development of such systems over the past two decades has been substantial, producing approaches ranging from classical collaborative filtering (Adomavicius & Tuzhilin, 2005) to deep neural networks capable of modeling complex sequential purchasing behavior (Zhang, Yao, Sun & Tay, 2019).

Despite this progress, a fundamental tension in grocery recommendation remains largely unaddressed. Standard recommender systems are designed to maximize the relevance of recommendations for the customer (Jannach & Bauer, 2020). In doing so, they treat every product as equally desirable from the retailer's perspective, recommending whichever items are most likely to be purchased regardless of their margin, inventory status, or strategic importance to the business. A system that consistently recommends low-margin commodity staples because customers reliably buy them performs well by conventional metrics while potentially failing the retailer on every dimension that determines commercial sustainability.

This issue is particularly acute in grocery retail for several reasons. First, the perishability of a substantial share of grocery products means that unsold inventory does not simply sit on a shelf but expires, generating direct financial losses (Amorim, Meyr, Almeder & Almada-Lobo, 2013). Second, the low-margin, high-turnover economics of the industry mean that small shifts in which products are recommended can have meaningful effects on profitability at scale (Ailawadi & Farris, 2017). Third, the basket-oriented nature of grocery shopping, where customers purchase multiple items in a single visit, means that recommendations influence not only individual product sales but the composition of entire baskets and the cross-selling opportunities they contain (Li, Dias, Jarman, El-Deredy & Lisboa, 2009).

The business case for making recommender systems aware of these commercial realities is therefore clear. However, how this can be achieved without undermining the customer utility

that makes recommender systems valuable in the first place is less well understood. A recommendation system that aggressively promotes high-margin or near-expiry products at the expense of relevance will erode customer trust and cease to function as an effective tool. The challenge is to identify approaches that serve both objectives simultaneously, or at least to characterize the trade-offs between them well enough to support informed operational choices. This challenge falls within the broader research agenda of multistakeholder recommender systems (Abdollahpouri et al., 2020), which explicitly recognizes that recommendation platforms operate within ecosystems where retailers, customers, and suppliers all have legitimate but potentially competing interests. While this agenda has produced a growing body of theoretical frameworks and isolated implementations (Jannach & Abdollahpouri, 2023), systematic empirical studies of business-aware recommendation in the specific context of grocery retail remain scarce. No existing work simultaneously addresses business awareness, next-basket prediction, and the repeat-consumption structure of grocery retail, nor does it offer a systematic comparison of how different integration methods perform under these conditions. This thesis fills that gap by developing and evaluating a framework that jointly optimizes customer utility and retailer business value within the specific economic constraints of grocery retail.

1.2 Research Questions

This thesis addresses the challenge of business-aware recommendation in grocery retail through the following research questions:

RQ1: How can standard grocery recommender systems be extended to incorporate business objectives without substantially compromising recommendation quality for customers?

RQ2: Which integration approach, post-processing reranking or model-level adjustment, achieves a more favorable trade-off between customer utility and business value?

RQ3: How do the structural properties of a recommender system influence its responsiveness to business-aware optimization?

RQ1 is the central question and motivates the overall experimental design. It builds on the observation by Jannach und Adomavicius (2017) that price and profit awareness in recommender systems can be achieved with only marginal losses in recommendation accuracy, and extends this line of inquiry to the multi-dimensional business context of grocery retail. RQ2 addresses the practical question of implementation strategy. Whether adjusting the outputs of existing models through reranking (Abdollahpouri, Burke & Mobasher, 2019) is sufficient, or whether business awareness needs to be incorporated into the learning process itself through model-based approaches (Rodriguez, Posse & Zhang, 2012). RQ3 emerged from the empirical results and addresses a finding with broader implications. The architectural choices made when designing a recommender system constrain how effectively it can be adapted to serve business objectives, a dimension of recommender system design that has received little attention in the literature (Jannach & Abdollahpouri, 2023).

1.3 Approach and Contributions

To address these questions, this thesis develops and evaluates a systematic framework for transforming standard recommender systems into business-aware systems within the grocery retail context. The framework is built around two established recommender models that represent different points on the complexity spectrum. TIFU-KNN (Hu, He, Gao & Zhang, 2020), a lightweight collaborative filtering approach based on temporal item frequency, and ReCANet (Ariannezhad et al., 2022), a neural network model designed specifically for repeat consumption prediction in grocery settings. Both models are first evaluated in their unmodified form to establish performance baselines, and then subjected to three distinct business-aware integration methods of increasing sophistication.

The first integration method is multiplicative scoring, a simple post-processing approach that reranks recommendations by multiplying predicted purchase probabilities with a business value score, following the general re-ranking paradigm described by Abdollahpouri et al. (2019). The second is Mixed Integer Programming, which frames reranking as a formal multi-objective optimization problem (Sürer, Burke & Malthouse, 2018) and evaluates two formulations. A weighted sum method and an epsilon-constraint method. The third consists of direct model adjustments, where business awareness is integrated into the recommendation model itself rather than applied afterward. An approach in line with model-based approaches to multistakeholder optimization (Rodriguez et al., 2012).

Central to the framework is a business value score, a composite metric that aggregates five dimensions of product-level commercial importance: contribution margin, inventory pressure, temporal urgency, cross-selling potential (Brijs, Swinnen, Vanhoof & Wets, 1999), and strategic priority. Since the Dunnhumby transaction dataset used in the experiments does not include operational business metrics such as product costs or inventory levels, these are generated through a simulation approach, following the precedent established by Sürer et al. (2018) and Jannach und Adomavicius (2017) in related multistakeholder settings. The business value score is deliberately designed as a flexible and extensible tool rather than a precise accounting instrument. Its purpose is to provide a comparable metric of business utility that can be systematically varied and evaluated alongside standard recommendation accuracy metrics. The precise calibration of such a score for any specific retailer is an open research question and a natural direction for future work.

The contributions of this thesis are threefold. First, it introduces and evaluates a structured framework for business-aware recommendation in grocery retail, providing a systematic comparison of integration methods across two fundamentally different model architectures. Second, it demonstrates empirically that meaningful improvements in business value can be achieved with modest accuracy trade-offs, particularly through MIP-based reranking applied to models with diverse candidate sets. Third, it identifies and explains a structural relationship between recommender system architecture and business-aware optimization responsiveness, showing that candidate set diversity and score distribution shape determine how effectively post-processing methods can improve commercial outcomes.

This thesis is positioned as a proof-of-concept study. The methods evaluated here demonstrate the feasibility of business-aware grocery recommendation and characterize the trade-off landscape in detail, but they do not constitute a production-ready system. Several compo-

nents, including the business value score calibration, the integration of real-time inventory data, and the computational scaling of MIP-based approaches to full catalog sizes, would require further engineering work before deployment. The primary contribution is a rigorous empirical understanding of what is possible and at what cost, which provides a principled foundation for subsequent applied development.

1.4 Thesis Structure

This thesis is organized as follows.

Chapter 2 reviews three areas of relevant literature. First, it traces the development of recommender systems from foundational collaborative filtering (Adomavicius & Tuzhilin, 2005; Burke, 2002) to modern deep learning architectures (Zhang et al., 2019; He et al., 2017). Second, it examines the distinctive characteristics of grocery retail that shape recommendation design (Sano, Machino, Yada & Suzuki, 2015; Ariannezhad et al., 2022). Third, it surveys multi-objective and multistakeholder recommender systems (Jannach & Abdollahpouri, 2023; Abdollahpouri et al., 2020), which directly inform the framework developed here.

Chapter 3 presents the foundational methodological components. It introduces the two base recommender models (TIFU-KNN and ReCANet), the Dunnhumby dataset, and the business value score and its construction.

Chapter 4 presents the business-awareness integration methods. It details the three approaches for incorporating business objectives, multiplicative scoring, Mixed Integer Programming based reranking, and direct model adjustments for both TIFU-KNN and ReCANet, and describes the evaluation framework, including the evaluation protocol, customer and business utility metrics, and trade-off analysis.

Chapter 5 presents and discusses the experimental results across all methods and both models. It begins with the unmodified baseline and proceeds through multiplicative scoring, MIP-based reranking, and direct model adjustments. Each section reports quantitative results and interprets them in the context of the research questions. The chapter closes with a customer-level impact analysis comparing the effects of business-awareness integration across user segments and models. Discussion is developed alongside the results rather than deferred to a separate chapter.

Chapter 6 reflects on the study's limitations and outlines directions for future research.

Chapter 7 summarizes the main findings of the thesis.

2 Literature Review

This thesis develops a business-aware recommender system for grocery retail. Three bodies of literature are relevant: recommender systems and their evolution toward multi-objective formulations, the operational characteristics of grocery retail that shape system requirements, and multi-objective and multistakeholder recommendation frameworks that directly inform the methodology. These are reviewed in Sections 2.1, 2.2, and 2.3, respectively. Section 2.4 positions this thesis within the reviewed literature and identifies the gaps it addresses.

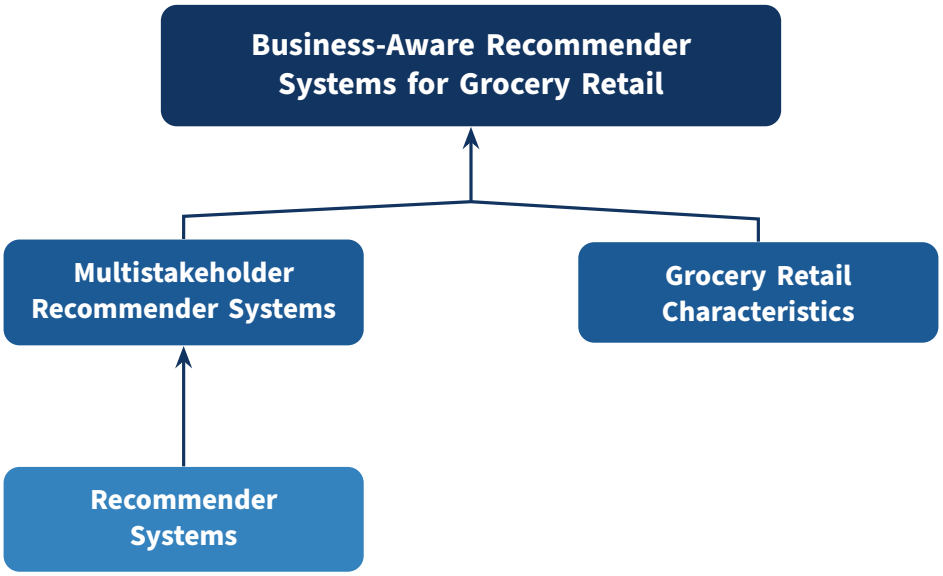

Figure 2.1: Literature review overview: Recommender Systems form the methodological foundation for Multistakeholder Recommender Systems, which together with Grocery Retail Characteristics inform the development of Business-Aware Recommender Systems for Grocery Retail.

2.1 Recommender Systems

As product catalogs have grown to an unmanageable scale, recommender systems have become essential tools for managing information overload. By leveraging user interaction data, item characteristics, and contextual signals, these systems surface personalized suggestions and direct users toward items matching their needs and preferences (Roy & Dutta, 2022). Applications range from e-commerce platforms such as Amazon and Netflix to social media and increasingly physical retail environments. The foundational challenge is the paradox of choice.

While extensive catalogs benefit users in principle, the sheer volume of options can impair decision-making. Recommender systems address this by filtering irrelevant offerings (Burke, 2002), thereby improving user experience while enabling businesses to increase engagement, retention, and revenue.

2.1.1 Foundational Approaches

Two principles underlie most recommender systems: collaborative filtering and content-based filtering, each modeling customer preferences through a distinct mechanism (Adomavicius & Tuzhilin, 2005).

Collaborative filtering operates on the assumption that users sharing similar past preferences will exhibit similar future interests (Roy & Dutta, 2022; Zhang et al., 2019). Recommendations are generated from collective user behavior without requiring explicit knowledge of item features. The approach divides into memory-based methods, which use the user-item rating matrix directly to identify similar users or items, and model-based methods, which learn latent factor representations through techniques such as matrix factorization (Koren, 2008). Despite its broad success, collaborative filtering faces well-documented challenges. The *cold-start problem* limits its usefulness for new users or items with sparse histories, *data sparsity* is pervasive when users rate only small fractions of available items, and analyzing shared behavioral data raises privacy concerns (Roy & Dutta, 2022).

Content-based filtering recommends items by measuring similarity between item attributes and a user's historical preferences, without reference to other users' behavior (Roy & Dutta, 2022). A movie recommender, for instance, would analyze genre, director, and cast to identify films similar to previously enjoyed ones. This approach avoids the cold-start problem for new items and is less privacy-sensitive than collaborative methods. Its primary limitations are reduced diversity (systems tend to recommend items similar to those already known, creating *filter bubbles*) and the considerable feature engineering effort required.

Hybrid approaches combine both techniques to mitigate their respective weaknesses (Burke, 2002; Roy & Dutta, 2022). Common strategies include weighted score combinations, context-dependent switching between methods, and feature augmentation. In practice, hybrid systems consistently outperform either approach alone. Content-based filtering handles cold-start cases while collaborative filtering provides diversity and accuracy once sufficient interaction history exists.

2.1.2 Deep Learning in Recommender Systems

Deep learning has substantially extended the capabilities of recommender systems. Neural networks capture non-linear interaction patterns between user and item representations that linear models cannot express (Roy & Dutta, 2022; Zhang et al., 2019), and they accommodate heterogeneous input types, including text, images, audio, and temporal sequences, within a unified architecture.

Several architectures have proven particularly valuable. Neural Collaborative Filtering replaces the inner product of traditional matrix factorization with multilayer neural networks, en-

abling the model to learn arbitrary interaction functions between user and item embeddings (He et al., 2017). Recurrent Neural Networks and their variants (LSTM, GRU) capture sequential dependencies and temporal dynamics in user behavior, which is especially relevant for session-based recommendation (Zhang et al., 2019). Convolutional Neural Networks extract features directly from raw multimedia content, extending content-based filtering beyond structured metadata. Autoencoders, particularly Variational Autoencoders, learn compressed preference representations from sparse rating data (Liang, Krishnan, Hoffman & Jebara, 2018).

2.1.3 Challenges and Research Frontiers

Several challenges persist despite these advances. Cold-start remains difficult for entirely new users or items, even when hybrid methods are applied. Scalability is non-trivial. Real-time systems serving millions of users require careful architectural design, and the computational demands of deep learning have motivated research into model compression and hardware acceleration. Fairness has emerged as a concern of increasing importance, as systems trained on historical data can amplify existing biases and produce discriminatory outcomes. Fairness-aware recommendation aims to balance predictive accuracy with equitable treatment (Ekstrand, Burke & Diaz, 2019).

2.1.4 Beyond Customer Utility

The standard formulation of recommendation as a user utility maximization problem is insufficient for commercial deployment. Real systems must also serve business objectives, including revenue, inventory management, and margin, that are not captured by accuracy metrics. This has motivated the field of multi-objective recommender systems, which explicitly optimize several objectives simultaneously rather than treating accuracy as a universal proxy. In grocery retail, where a single recommendation must simultaneously serve customer preferences, retailer profitability, inventory constraints, and supplier relationships, the multi-objective framing is not optional. It is the only framing that corresponds to the actual decision problem.

2.2 Characteristics of the Grocery Retail Industry

Grocery retail differs from the domains in which most recommendation research has been conducted, such as movies, books, and music, and these differences are consequential for system design (Sano et al., 2015; Li et al., 2009). The characteristics described in this section are not minor variations on a standard e-commerce setting. They change what a recommendation system needs to optimize and how its performance should be evaluated.

2.2.1 Repeat Consumption and High Purchase Frequency

Unlike most retail contexts where purchases are irregular or one-time events, grocery shopping is characterized by repetitive consumption and high purchase frequency (Bhagat, Muralidharan, Lobzhanidze & Vishwanath, 2018). Consumers regularly repurchase the same or similar products to meet ongoing household needs, a pattern termed *reoccurring consumption* (Ariannezhad et al., 2022). This creates temporal structure in purchase data as the likelihood of buying a product is often tied directly to how long ago it was last purchased (Guidotti et al., 2017; Anderson, Kumar, Tomkins & Vassilvitskii, 2014).

Purchase cycles vary considerably by category. Staples such as milk or bread may be replenished weekly or more frequently, while cleaning supplies may follow monthly or quarterly cycles (Guidotti et al., 2019). A recommendation model that ignores this temporal structure will misfire systematically, either pushing products the customer just bought or missing the window when they are actually running low. This is a direct challenge to collaborative filtering approaches developed for the movie or book domain, where the recommendation objective is primarily discovery. In grocery, discovery is secondary; most value comes from reliably surfacing the right replenishment items at the right time (Li et al., 2009; Ariannezhad et al., 2022). Accurately modeling this requires temporal awareness that standard approaches lack (Guidotti et al., 2017). The high transaction frequency does, however, provide a data richness that makes detailed preference modeling feasible, provided the system is designed to exploit it (Sano et al., 2015).

2.2.2 Product Perishability and Temporal Constraints

A substantial fraction of grocery products have limited shelf lives. Fresh produce, dairy, meat, and bakery items all expire, and their value drops to zero (or below, once disposal costs are counted) if unsold before their expiration date (Lau, Nakandala & Shum, 2016). For the retailer, this creates a persistent inventory management tension as overstocking leads to waste and write-offs, while understocking means lost sales and reduced customer satisfaction (Herbon, Levner & Cheng, 2014).

Perishability introduces a dimension largely absent in other recommendation domains. A system optimizing purely for preference will ignore the retailer's need to move items approaching expiration. Business-aware systems must therefore incorporate remaining shelf life, inventory turnover rates, and seasonal availability, either as hard constraints or as objectives to be traded off against relevance (Amorim et al., 2013; Rijpkema, Rossi & van der Vorst, 2014).

2.2.3 Physical Store Environment and Spatial Constraints

The majority of grocery transactions still occur in physical stores, and even online orders are typically fulfilled from physical shelves (Sano et al., 2015). This creates constraints absent from purely digital retail. In-store, shoppers follow a physical path; product placement, store layout, and the effort of carrying items all influence purchasing behavior. The tangibility of products allows sensory evaluation. Customers can assess freshness directly and make decisions based on visual cues or in-store promotions (Christodoulou et al., 2017).

Delivering recommendations in a physical environment requires different interfaces than a website, such as mobile apps, digital shopping lists, and shelf displays, and the recommendations themselves should respect the logic of a physical shopping trip (Sano et al., 2015; Christodoulou et al., 2017). For online channels, the challenges shift. Minimum order values, delivery slot availability, and out-of-stock substitution all complicate the recommendation problem. The physical fulfillment process carries real costs that can reasonably inform recommendation strategy (Jansen, Bennin, van Kleef & Van Loo, 2024).

2.2.4 Basket-Oriented Purchasing and Complementarity

Grocery shopping is inherently a basket activity. Consumers purchase multiple items per visit to serve different needs, and these items are often related by recipe, meal occasion, or consumption habit (Li et al., 2009). The resulting co-purchase patterns, such as pasta with sauce or coffee with milk, are strong, consistent, and commercially relevant (Brijs et al., 1999).

This basket structure has direct implications for recommendation design. Item-level suggestions remain useful, but basket-level recommendations that propose complementary products or complete consumption solutions can deliver more value to both the customer and the retailer (Le, Lauw & Fang, 2019). Market basket analysis and association rule mining are natural methodological tools here, complementing collaborative filtering with explicit co-purchase signals (Kutuzova & Melnik, 2018). From a business perspective, the basket also provides a natural unit for margin optimization: a higher-margin complementary item recommended alongside a staple is a low-friction intervention that is transparent in value to the customer yet meaningful to retailer profitability (Park et al., 2018).

2.2.5 Low-Margin/High-Turnover Business Model

Grocery retail operates on some of the thinnest profit margins in any retail sector. Net margins typically fall between 1% and 3%, frequently below 2% (Trejo-Pech & White, 2024). Profitability depends on volume and operational efficiency rather than per-unit economics. Inventory turns rapidly and even faster in fresh categories, because capital cannot be left idle on shelves when margins are this thin.

Slow-moving inventory is doubly costly as direct carrying costs accumulate, and the tied-up capital is unavailable for more productive uses. Efficient retailers minimize days of inventory outstanding and manage their cash conversion cycle tightly. For a recommender system, this implies that there is concrete financial value in shifting inventory that would otherwise stagnate.

A system designed only to maximize relevance will not capture this value. One designed to balance relevance with inventory economics can (Ailawadi & Farris, 2017).

2.2.6 Price Sensitivity and Promotional Intensity

Given that groceries represent a recurring, substantial portion of household budgets and that products are relatively standardized across retailers, consumers are highly price-sensitive. The result is an industry characterized by intense promotional activity, including weekly deals, temporary price reductions, and loyalty programs, that creates a dynamic pricing environment in which the relative attractiveness of products changes continuously (Ailawadi & Farris, 2017; Krafft & Mantrala, 2010).

For recommender systems, promotions present both an opportunity and a risk. Aligning suggestions with active deals can increase relevance and conversion rates (Dhar, Morrison & Raju, 2001). Systems that over-index on promoted items, however, risk training customers to purchase only on deal, gradually eroding margins and displacing preference-based recommendations. Business-aware systems must balance these competing effects. Using promotional signals where they genuinely improve recommendations without allowing them to crowd out the underlying preference model.

2.3 Multi-Objective and Multistakeholder Recommender Systems

The traditional recommender systems paradigm optimizes a single objective. Maximizing relevance or prediction accuracy for the end user. This user-centric formulation has driven substantial methodological progress, but it is an incomplete model of real-world deployment, where multiple parties have stakes in the recommendation outcome and accuracy alone captures neither business realities nor fairness considerations. The recognition of this gap has motivated multi-objective recommender systems, and within this field, multistakeholder recommendation as a framework that explicitly accounts for the competing interests of all parties involved.

2.3.1 Defining Multi-Objective Recommender Systems

A multi-objective recommender system optimizes or balances more than one goal simultaneously (Jannach & Abdollahpouri, 2023). Rather than collapsing all quality dimensions into a single accuracy metric, these systems treat recommendation quality as inherently multi-dimensional. Five primary categories of objectives have been identified in the literature (Jannach & Abdollahpouri, 2023; Zheng et al., 2021):

1. **Quality objectives**: Multiple dimensions of recommendation quality for users, including relevance, diversity, novelty, serendipity, and coverage. These frequently trade off against each other; optimizing for accuracy tends to favor popular items, which reduces novelty and diversity.
2. **Multistakeholder objectives**: The interests of multiple parties, including consumers, service providers, item suppliers, and potentially broader societal actors. These are frequently in tension; maximizing platform revenue does not automatically align with maximizing user satisfaction.
3. **Long-term versus short-term objectives**: The tension between immediate metrics such as click-through rates and longer-term goals including customer lifetime value, retention, and trust. Short-term optimization can erode the foundation of sustainable customer relationships.
4. **User interface objectives**: Trade-offs related to how recommendations are presented, including cognitive load, information density, and the balance between simplicity and comprehensiveness.
5. **Engineering objectives**: System-level considerations including computational efficiency, scalability, explainability, and maintainability.

The practical case for multi-objective optimization is straightforward. A system recommending only popular items may achieve strong accuracy on historical data while failing to expose long-tail items, providing little discovery value, and gradually homogenizing consumption in ways that reduce platform value over time (Jannach & Bauer, 2020). Accuracy as a sole objective is a proxy that regularly misfires.

2.3.2 Multistakeholder Recommender Systems

Multistakeholder recommendation explicitly recognizes that recommender systems operate within ecosystems where multiple parties have legitimate interests that may align or conflict (Abdollahpouri et al., 2020). The framework asks whose utility the system should optimize, and how conflicts should be resolved.

Stakeholder Identification. Three categories of stakeholders are typically distinguished (Abdollahpouri et al., 2020; Abdollahpouri & Burke, 2022). Consumers receive and act on recommendations; their interests center on relevance, discovery, privacy, and transparency. Providers, such as retailers, platforms, and streaming services, operate the system and hold multi-faceted objectives including revenue, user retention, and inventory efficiency. Platforms and society constitute a third category whose relevance varies by context; in grocery retail, this may include public health authorities concerned with nutrition, environmental stakeholders focused on food waste, and regulatory bodies overseeing fair market practices.

The tension between consumer and retailer perspectives is central to this thesis. From the consumer side, recommendations should balance routine replenishment with relevant discovery, respect budget constraints, and reduce shopping effort. From the retailer side, the same recommendations should also promote higher-margin items, move perishable inventory before expiration, and support promotional strategy. These objectives are not inherently opposed, but neither are they automatically aligned. A system that ignores the tension does not resolve it.

2.3.3 Implementations of Multistakeholder Recommender Systems

Several implementations demonstrate that multistakeholder objectives can be integrated into practical systems without sacrificing recommendation quality.

Sponsored Recommendations for Online Grocery Retail. Malthouse et al. (2019) developed a multistakeholder framework for online grocery that balances advertising revenue against user utility. By weighting these two objectives, the model identifies integration strategies for sponsored content that avoid significant degradation of recommendation quality. Their key finding is that substantial advertising revenue gains are achievable at a small cost to relevance. The approach functions as a modular post-processing step, making it straightforward to integrate into existing systems.

Price and Profit-Aware Recommendation. Jannach und Adomavicius (2017) examined how recommender systems can balance user relevance with business profitability. Their results show that platforms can optimize for profit through hybrid scoring and mathematical programming while accepting only marginal losses in recommendation accuracy. Importantly, the authors argue for a long-term orientation as sustainable gains depend on maintaining user trust rather than extracting maximum short-term transactional revenue, a finding directly relevant to the grocery context.

Multistakeholder Recommendation with Provider Constraints. Sürer et al. (2018) addressed provider fairness in multistakeholder marketplaces, introducing constraints that guarantee minimum exposure for each provider to prevent recommendation algorithms from systematically favoring dominant suppliers. Their results show that these guarantees can be enforced with minimal impact on user relevance, as highly rated alternatives from underrepresented providers often serve as near-equivalent substitutes for users.

Learning to Rank for Multistakeholder Optimization. Nguyen et al. (2017) proposed a learning-to-rank framework for hotel recommendations that integrates traveler preferences, hotel margins, and platform revenue into a unified two-stage model trained on real booking outcomes. Live A/B testing demonstrated improvements in both conversion and profitability, confirming that jointly modeling stakeholder interests through a data-driven approach can outperform systems that treat business objectives as secondary constraints.

2.3.4 Approaches to Solving Multistakeholder Recommendation Problems

Re-ranking Approaches. Re-ranking is the most widely adopted strategy for multistakeholder optimization, largely because of its modularity. A base recommender generates an initial candidate set ranked by relevance, and a post-processing step reorders the list to incorporate additional objectives (Abdollahpouri et al., 2019; Jugovac, Jannach & Lerche, 2017). Two variants dominate the literature. Greedy re-ranking selects items sequentially, at each step choosing the candidate that maximizes a combined objective given already selected items (Adomavicius & Kwon, 2012). This is computationally efficient, but provides no guarantee of global optimality. Optimization-based re-ranking formulates the problem as mathematical programming, typically integer linear programming, to find near-optimal solutions subject to hard constraints (Sürer et al., 2018). This produces better solutions at higher computational cost and limited scalability to very large item catalogs.

Multi-Objective Optimization Approaches. Multi-objective optimization (MOO) explicitly maintains separate objectives and characterizes the set of Pareto-optimal solutions, configurations where no objective can be improved without worsening another (Jannach & Abdollahpouri, 2023). Rather than producing a single recommendation list, these approaches generate a frontier of trade-off solutions, allowing decision-makers to choose an operating point based on business priorities. The cost is computational. Generating and evaluating multiple Pareto solutions is expensive, and real-time deployment under latency constraints may render this infeasible. Selecting among Pareto solutions also requires additional decision-making infrastructure.

Model-Based Approaches. Rather than post-processing existing recommendations, model-based approaches integrate multistakeholder objectives directly into the learning procedure

(Rodriguez et al., 2012). Three paradigms have been explored. Multi-task learning trains a single model to simultaneously predict multiple outputs, namely user ratings, purchase probability, and profit margin, sharing representations across tasks while maintaining task-specific output layers. Adversarial training introduces a component that attempts to maximize one objective while the main model balances multiple competing goals. Reinforcement learning formulations model recommendation as sequential decision-making, where the agent receives combined rewards reflecting multiple stakeholder utilities and learns a policy that balances immediate and long-term value.

Model-based approaches can in principle learn more nuanced trade-offs than post-processing allows, since the learned representations adapt to the multi-objective setting from the outset. In practice, they require retraining when business priorities shift, lack the modularity of re-ranking approaches, and demand more complex training procedures and larger datasets to learn multi-objective trade-offs reliably.

2.4 Conclusion and Positioning of this Thesis

The reviewed literature establishes that recommendation algorithms can be designed to serve multiple parties with competing interests, and that doing so is both technically feasible and commercially valuable (Jannach & Abdollahpouri, 2023; Abdollahpouri et al., 2020).

As shown in Table 2.1, no existing work simultaneously addresses business awareness, next-basket prediction, the repeat-consumption structure of grocery retail, and a comparative analysis of integration methods. This thesis addresses all four. Drawing on the grocery retail characteristics discussed in Section 2.2, specifically the low-margin/high-turnover economics (Ailawadi & Farris, 2017), perishability constraints (Amorim et al., 2013), and basket-oriented purchasing behavior (Li et al., 2009), a system is developed that optimizes customer utility and retailer business value jointly, rather than treating one as a constraint on the other.

The specific contributions are: (1) adaptation of multistakeholder optimization frameworks (Abdollahpouri et al., 2020) to the constraints of grocery retail; (2) integration of inventory management and perishability considerations into the recommendation objective; (3) empirical evaluation on real-world grocery transaction data quantifying the trade-offs between customer utility and business metrics; and (4) a comparative analysis of integration methods demonstrating that carefully designed multistakeholder systems improve outcomes for both customers and retailers relative to single-objective baselines. The design, implementation, and evaluation of this system are detailed in the following chapters.

Reference	Description	Business Awareness	Next Basket	Grocery / Repeat	Integration Comp.
Jannach und Adomavicius (2017)	Price and profit-aware recommendation balancing user relevance and business	✓	×	×	×
Sürer et al. (2018)	Multistakeholder recommendation with provider fairness constraints	✓	×	×	×
Nguyen et al. (2017)	Learning-to-rank framework balancing user preference and hotel margins	✓	×	×	×
Rodriguez et al. (2012)	Model-based approaches for multiple objective optimization	✓	×	×	×
Abdollahpouri et al. (2019)	Post-processing re-ranking to manage popularity bias	✓	×	×	×
Malthouse et al. (2019)	Allocating sponsored recommendations in online grocery	✓	×	✓	×
Li et al. (2009)	Basket-sensitive random walk for grocery shopping	×	✓	✓	×
Bhagat et al. (2018)	Modeling repeat purchase recommendations	×	×	✓	×
Guidotti et al. (2019)	Temporal annotated recurring sequences for market basket prediction	×	✓	✓	×
Ariannezhad et al. (2022)	ReCANet model for repeat consumption in next basket grocery	×	✓	✓	×
This Thesis (2026)	**Multi-objective next basket optimization comparing integration methods**	✓	✓	✓	✓

Table 2.1: Literature review matrix: Comparison of related work based on key characteristics.

3 Methodology

Standard recommender systems are designed to maximize customer satisfaction by predicting and suggesting products that align with individual user preferences. While effective at improving user experience, such customer-centric approaches typically disregard operational business objectives, including profit margins and inventory management. To address this limitation, we develop and evaluate three methods that extend existing, single-objective recommender systems into multi-objective frameworks capable of jointly optimizing customer utility and business value.

The central idea is that recommendations can simultaneously account for business metrics, such as product margins or remaining shelf life, alongside customer utility criteria. A business-aware system might, for instance, give priority to high-margin products or items approaching expiration, while preserving recommendation relevance for the customer. Balancing these two objectives poses methodological challenges but offers considerable practical value in commercial settings.

To investigate this trade-off systematically, we evaluate three approaches to incorporating business awareness into existing recommendation algorithms. Multiplicative Scoring adjusts item scores after the ranking and reorders accordingly. Mixed Integer Programming reformulates the selection of the final recommendation list as an optimization problem that maximizes a combination of customer and business utility. Model Adjustment integrates business information directly into the training objective of the underlying recommendation model. These three approaches represent different points along the spectrum of implementation complexity, computational overhead, and depth of integration between the two objectives.

The resulting business-aware systems are then evaluated against their unmodified counterparts using a dual-metric framework. Customer utility is assessed through standard quality metrics, namely recall and normalized discounted cumulative gain (NDCG). Business utility is captured through a composite business value score that aggregates product-level operational features. This combined evaluation allows for a structured analysis of the accuracy-business utility trade-off and supports identifying which integration approach best balances the two competing objectives. Figure 3.1 illustrates the overall structure of the proposed framework.

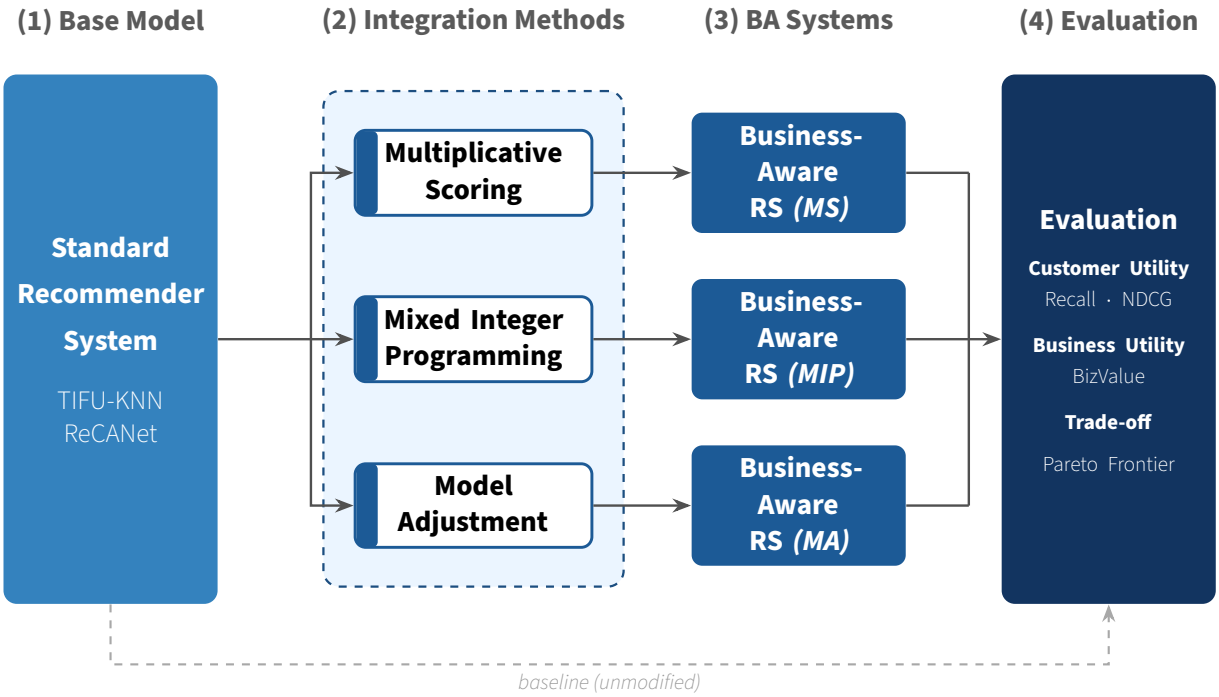

Figure 3.1: Framework overview: the standard recommender system is extended via three integration methods to produce business-aware systems, which are evaluated against the unmodified baseline across customer utility and business value metrics.

3.1 Notation and Terminology

To ensure mathematical consistency across the distinct predictive models and the business-awareness adjustments, this thesis adopts a unified notation framework. Tables 3.1-3.2 summarize the primary variables, sets, and parameters used throughout the methodology.

Symbol	Description
General Sets and Indices	
U	Set of all users
u	A specific target user, where $u \in U$
I	Set of all available items in the catalog
i	A specific target item, where $i \in I$
k	Cardinality of the final recommendation list
C	Size of the truncated candidate set prior to reranking
Predictions and Scores	
$\hat{p}_{u,i}$	Predicted purchase probability for user u and item i
b_i	Business Value Score (BVS) for item i
w_i	Normalized business weight for item i, derived from b_i to $[0, 1]$
$S_{u,i}$	Fused score for user u and item i
$\hat{R}_u^k$	Final recommended set of k items for user u
$y_{u,i}$	Binary label: 1 if i is in user u's target basket, 0 otherwise

Table 3.1: Unified notation: general sets, indices, predictions, and scores.

3.2 TIFU-KNN: Temporal Item Frequency-Based User KNN

TIFU-KNN is a k-nearest neighbors approach designed specifically for next-basket recommendation. Rather than learning complex patterns through neural networks, it takes an interpretable path by explicitly exploiting personalized item frequency information, combined with temporal dynamics and collaborative signals (Hu et al., 2020).

3.2.1 Personalized Item Frequency as a Predictive Signal

The method builds on the observation that personalized item frequency, how often a specific user has purchased each item, carries strong predictive information about future purchases. An analysis of real-world datasets by the original authors identifies two recurring patterns that motivate the design.

Repeated Purchase Pattern. Users tend to repurchase items they have bought before, and higher past purchase frequency correlates with a higher likelihood of future purchase. This pattern is particularly pronounced in grocery retail, where consumers regularly restock staple products.

Collaborative Purchase Pattern. Items that a given user purchases repeatedly are also likely to be purchased by similar users. This extends collaborative filtering principles to the frequency domain. Across four datasets, examining 100-300 nearest neighbors was sufficient to recover between 55% and 98% of items in the target basket, consistently outperforming the repeated purchase signal alone.

3.2.2 User Representation with Hierarchical Temporal Weighting

The central component of TIFU-KNN is a user representation vector that aggregates the purchase history while accounting for temporal dynamics. Given a user's historical baskets $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, where $\mathbf{v}_t$ is a binary vector indicating which items appeared in the t-th basket, the method applies two successive decay mechanisms.

Within-Group Weighting. The n baskets are first partitioned into m groups of approximately equal size, with $g = \lceil n/m \rceil$ baskets per group. Within each group, the j-th basket (ordered from oldest to newest) is weighted by $r_b^{(g-j)}$, where $r_b \in [0, 1]$ is the within-group decay ratio. The resulting group vector is:

$$\mathbf{v}_{\text{group}_l} = \frac{\sum_j r_b^{(g-j)} \cdot \mathbf{v}_j}{g} \tag{3.1}$$

Across-Group Weighting. The m group vectors are subsequently weighted by their recency. The l-th group vector receives weight $r_g^{(m-l)}$, where $r_g \in [0, 1]$ is the across-group decay ratio, yielding the personal user representation:

$$\mathbf{u}_{\text{pers}} = \frac{\sum_{l=1}^{m} r_g^{(m-l)} \cdot \mathbf{v}_{\text{group}_l}}{m} \tag{3.2}$$

This two-level structure reflects the nature of preference drift. Consecutive baskets tend to differ only marginally, whereas baskets separated by longer intervals may reflect meaningful shifts in purchasing behavior. The hierarchical decay captures both fine-grained short-term changes and longer-term trends within a single representation. Figure 3.2 illustrates the grouping and weighting procedure.

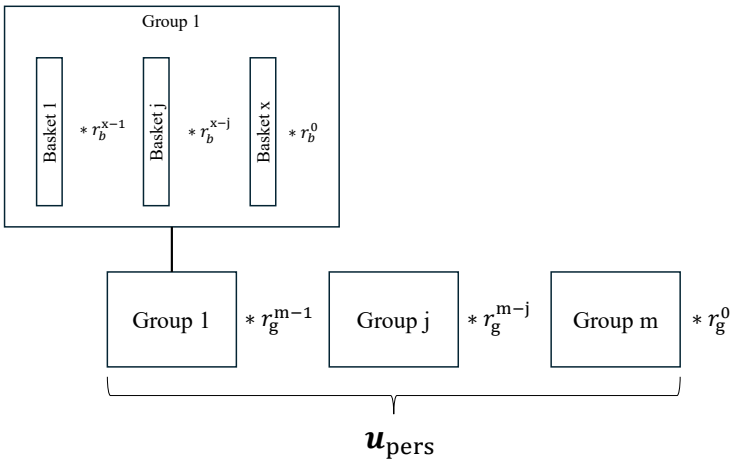

Figure 3.2: Purchase history partitioned into m groups. Temporal decay factors are applied both within and across groups to construct the personal user representation $\mathbf{u}_{\text{pers}}$.

3.2.3 Prediction

The final item score for a given user combines a personal and a collaborative component. The personal component $\mathbf{u}_{\text{pers}}$ encodes the user's own purchase frequencies weighted by recency, as described above. The collaborative component $\mathbf{u}_{\text{collab}}$ is the average representation vector across the k nearest neighbors in user space, capturing the collective purchase tendencies of similar users. The predicted score vector is:

$$\hat{p}_u = \alpha \cdot \mathbf{u}_{\text{pers}} + (1 - \alpha) \cdot \mathbf{u}_{\text{collab}} \tag{3.3}$$

where $\alpha \in [0, 1]$ controls the relative contribution of personal versus collaborative information. Items are ranked by their score in $\hat{p}_u$, and the top-k items form the recommended basket. Higher values of α place greater emphasis on the user's individual history, while lower values give more weight to neighborhood patterns.

3.2.4 Interpretability and Computational Properties

The contribution of each component in TIFU-KNN is directly interpretable. The temporal decay parameters r_b and r_g make explicit how much influence different periods of purchase history exert on the final score, and α transparently governs the personal-collaborative balance.

At inference time, the method is computationally lightweight. User representations can be computed offline, after which prediction reduces to nearest-neighbor lookup and weighted vec-

tor averaging. This stands in contrast to recurrent neural approaches, where inference requires a sequential forward pass through the network for each user.

3.3 ReCANet: Repeat Consumption-Aware Neural Network

ReCANet is a neural network model for next-basket recommendation that focuses exclusively on predicting repeat purchases, items that a user has previously bought and is likely to buy again. This design is motivated by empirical observations showing that repeat items, while constituting only about 1% of the total product catalog, account for over 54% of recommendation performance in grocery retail contexts (Ariannezhad et al., 2022). Rather than scoring all available items, ReCANet restricts its candidate set to each user's purchase history and predicts for each previously purchased item whether it will appear in the user's next basket.

3.3.1 Model Architecture

Input Representation. For each user-item pair, the model receives three inputs: the item identifier i, the user identifier u, and a history vector $\vec{h}$ that encodes the temporal pattern of past purchases.

The history vector is constructed for a given window size W by recording the W most recent occasions on which user u purchased item i and computing the inter-purchase gaps. Formally, let item i have been purchased by user u at basket positions $ba_1 < ba_2 < \cdots < ba_q$ within a total history of n baskets. The history vector $\vec{h} \in \mathbb{R}^W$ is defined as:

$$\vec{h}_j = \begin{cases} ba_{j+1} - ba_j & \text{for } j = 1, \ldots, W-1 \\ (n+1) - ba_W & \text{for } j = W \end{cases} \tag{3.4}$$

As an illustration, consider $n = 99, W = 5$, and item i purchased at baskets $\{45, 78, 95, 97, 99\}$. Taking the five most recent occurrences yields:

$$\vec{h} = [78 - 45,\ 95 - 78,\ 97 - 95,\ 99 - 97,\ 100 - 99] = [33,\ 17,\ 2,\ 2,\ 1] \tag{3.5}$$

Decreasing values toward the end of $\vec{h}$ indicate recent, frequent consumption and constitute a strong signal for repeat purchase prediction (Ariannezhad et al., 2022).

User and Item Embeddings. Users and items are each mapped to a dense vector space through separate embedding layers. This allows the network to capture latent similarity. Users with similar purchasing behavior and items that tend to be bought together become proximate in their respective embedding spaces. The two embedding vectors are concatenated and passed through a feed-forward layer with ReLU activation to produce a unified user-item representation.

Consumption Pattern Modeling. To incorporate temporal information, the user-item representation is replicated W times and each copy is concatenated with the corresponding element of $\vec{h}$. This produces a sequence of W vectors that combines user-item context with inter-purchase gap information at each time step. The sequence is then processed by a stack of two LSTM layers, which model the sequential dependencies in consumption patterns.

The LSTM output is passed to a two-layer feed-forward network with ReLU activation, followed by a sigmoid output layer that produces the purchase probability $\hat{p}_{u,i} \in [0, 1]$ for item i in user u's next basket.

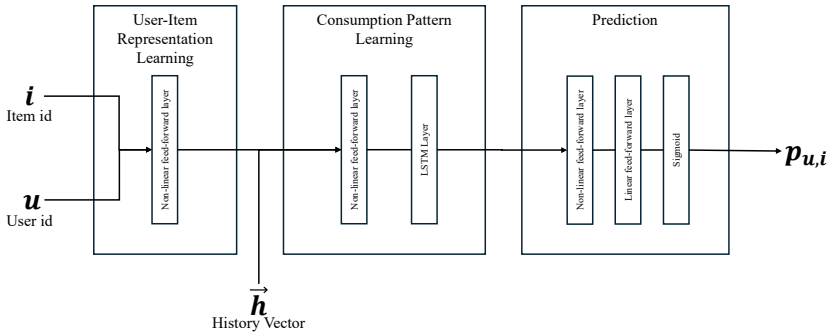

Figure 3.3: User and item embeddings are combined into a joint representation, which is concatenated with the history vector $\vec{h}$ and passed through a non-linear feed-forward layer and LSTM to capture temporal consumption patterns, before a final MLP with sigmoid activation outputs the probability p_x^u of item x appearing in the user's next basket.

3.3.2 Training

Sample Construction. Training samples are derived from each user's last L baskets. This truncation limits the influence of distant history, focusing the model on recent behavior, and prevents users with very long purchase records from disproportionately dominating gradient updates.

For each basket ba_t in a user's last L baskets, all items purchased in any of the preceding baskets $ba_1, \ldots, ba_{t-1}$ are treated as candidates. Each candidate yields one training sample, consisting of the item identifier, user identifier, and history vector. Items that appear in ba_t receive a positive label, all remaining candidates receive a negative label. The resulting dataset

is inherently imbalanced, which mirrors the actual distribution encountered at inference time.

Loss Function. The model is trained by minimizing the binary cross-entropy loss over all user-item training pairs $\mathcal{T}$:

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} [y_{u,i} \log(\hat{p}_{u,i}) + (1 - y_{u,i}) \log(1 - \hat{p}_{u,i})] \tag{3.6}$$

where $y_{u,i} = 1$ if item i appears in the target basket of user u and $y_{u,i} = 0$ otherwise. Parameters are updated via mini-batch stochastic gradient descent using the Adam optimizer.

3.3.3 Inference

At test time, the candidate set for each user consists of all items present in their training history. For each candidate, the model constructs the corresponding item identifier, user identifier, and history vector using the same procedure as during training, and outputs a purchase probability. Candidates are ranked by this probability, and the top-k items form the recommended basket.

Restricting the candidate set to previously purchased items provides a substantial computational advantage over full-catalog ranking. Rather than scoring potentially tens of thousands of products, ReCANet evaluates only the few hundred items each user has bought before, reducing inference cost without sacrificing recommendation quality.

3.4 Data

3.4.1 Dataset Overview

The experiments are conducted on the Dunnhumby transaction dataset, which comprises 2,595,732 transaction records from 2,500 unique households over a period of 711 days. The dataset contains 12 variables, including household identifiers, basket IDs, product IDs, quantities, sales values, and temporal information. It is a synthetic dataset constructed from patterns observed in real in-store data and contains no missing values or duplicates.

Customer behavior varies considerably across the dataset, as illustrated in Figure 3.4. Shopping frequency ranges from 1 to 1,300 trips per household, with a mean of 111 and a median of 79, indicating a right-skewed distribution in which a small number of highly active households account for a disproportionate share of activity. Total household spending ranges from €8.17 to €38,319.79, with a mean of €3,222.99 and a median of €2,157.75. Both distributions exhibit a pronounced right skew, with the mean sitting well above the median, reflecting a long tail of high-frequency, high-spending power users. Basket size averages 9.4 items with a median of 5, and basket value averages €29.14 with a median of €17.07, again right-skewed, driven by infrequent but large stock-up purchases. The consistent divergence between means and medians across all four dimensions confirms that population-level approaches would be inappropriate, and motivates the personalized modeling strategy adopted in this thesis.

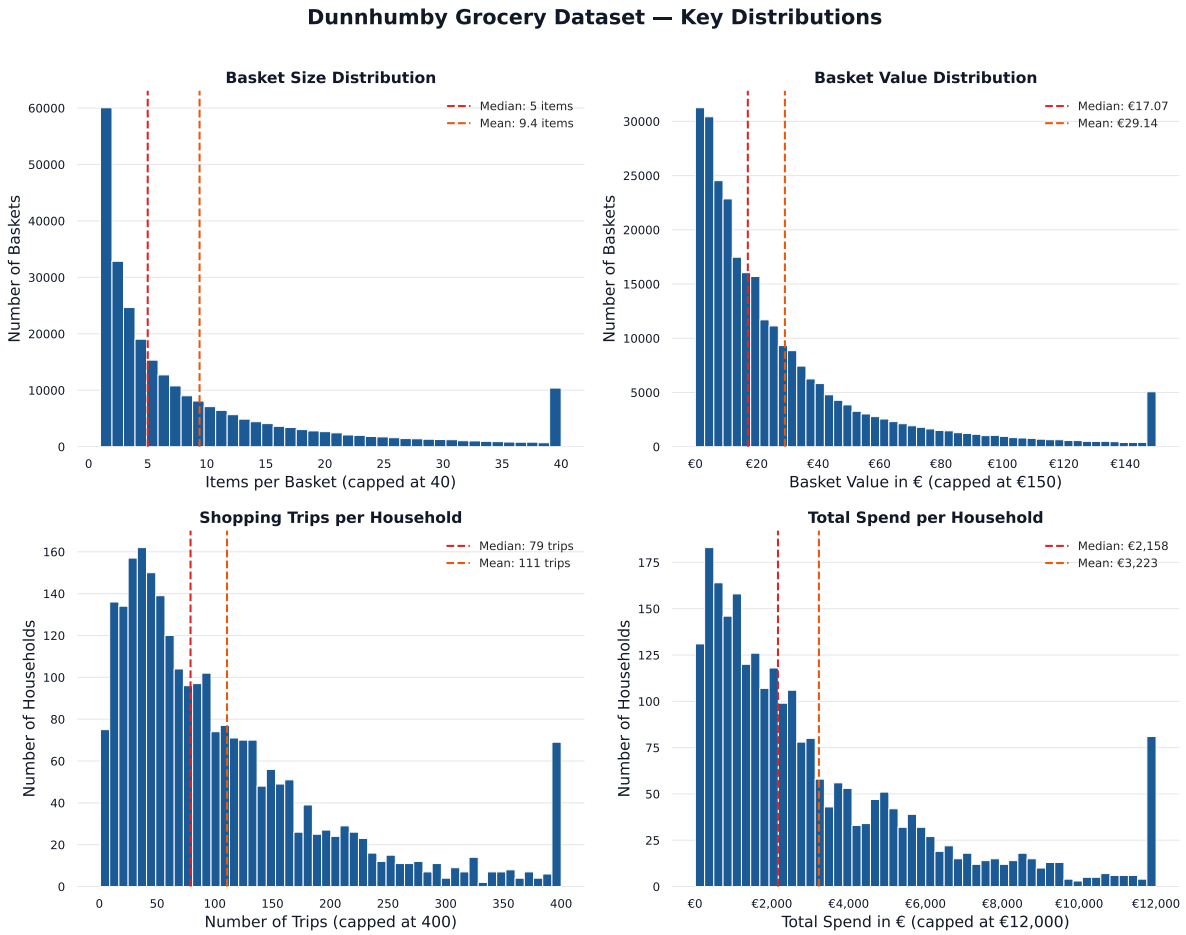

Figure 3.4: Key distributional characteristics of the Dunnhumby dataset. All four distributions exhibit right skew, with the mean (orange dashed line) consistently exceeding the median (red dashed line). Outliers are capped for readability: basket size at 40 items, basket value at €150, shopping trips at 400, and household spend at €12,000.

The product catalog consists of 92,339 unique items, reflecting the broad assortment typical of grocery retail. A Pareto analysis shows that approximately 10,679 products (11.6% of the catalog) account for 80% of all transactions, as illustrated in Figure 3.5.

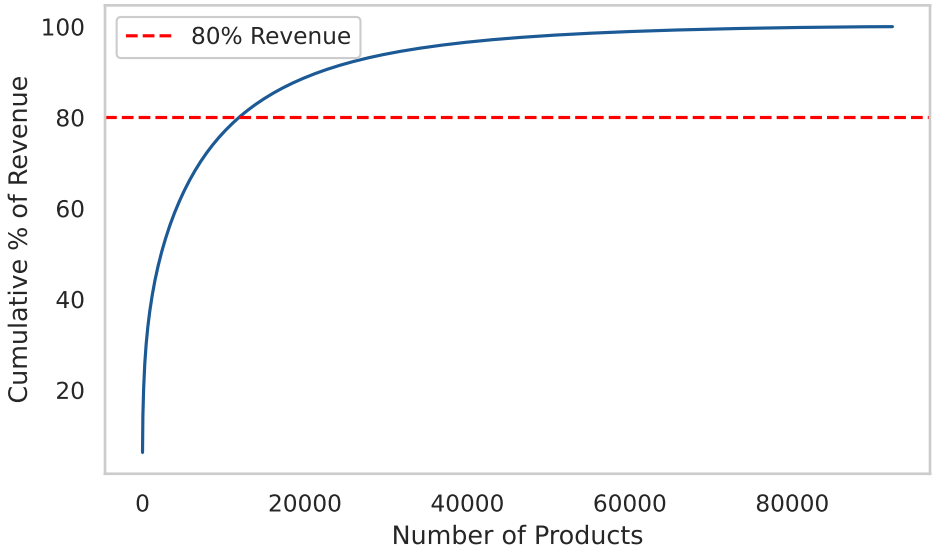

Figure 3.5: Cumulative share of total revenue by product, ordered from highest to lowest individual revenue contribution. The curve highlights the concentration of sales among a small fraction of the catalog.

3.4.2 Product Catalog Characteristics

The transaction dataset is complemented by a product catalog comprising 92,339 unique items organized across 7 departments, 308 commodity categories, and 2,383 sub-commodity categories, supplied by 6,476 distinct manufacturers. Each product is characterized by its department, brand type, and commodity description. This catalog serves as the informational basis for the business value score simulation described in Section 3.5, where commodity descriptions are used to assign products to margin and perishability clusters.

Figure 3.6 shows the distribution of total revenue across departments. `GROCERY` dominates with 55.5% of all revenue, followed by `DRUG & GM` at 14.3%, and `PRODUCE AND MEAT` each contributing approximately 7.6%. The remaining departments collectively account for less than 30% of revenue, confirming the low-margin, high-turnover structure characteristic of grocery retail discussed in Section 2.2. A small number of high-frequency commodity departments drive the overwhelming majority of commercial activity.

Figure 3.6 disaggregates this further to the commodity level. The fifteen highest-revenue commodity categories are dominated by everyday consumables. `SOFT DRINKS`, `FLUID MILK`, `COLD CEREAL`, `SOUP`, and `BAKED BREAD`, alongside fresh and frozen protein categories including `BEEF`, `CHICKEN`, `PORK`, `DELI MEATS`, and `FROZEN MEAT DINNERS`. This composition has direct implications for the BVS simulation. The protein and bakery categories are precisely those with the shortest repurchase cycles and the greatest perishability exposure. The category structure visible in Figure 3.6 thus provides the empirical grounding for the distributional assumptions adopted in Section 3.5.

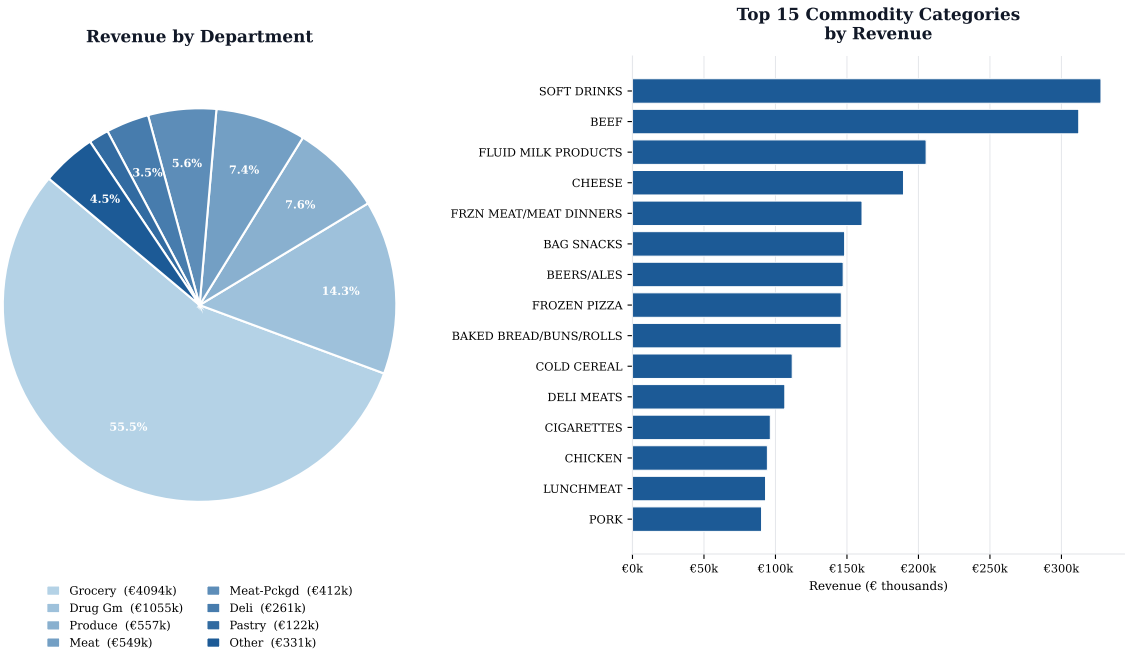

Figure 3.6: Left: revenue share by department, with Grocery accounting for over half of total revenue. Miscellaneous and kiosk categories are excluded. Right: the fifteen highest-revenue commodity categories, illustrating the dominance of everyday consumables and perishable protein categories.

3.4.3 Repeat Purchase Analysis

Definition. A repeat purchase is defined as any subsequent purchase of the same product by the same household following the initial purchase occasion. This definition targets product-specific repurchase behavior rather than general shopping frequency. For each household-product combination, the date of first purchase serves as the reference point, and all later purchases of that product by the same household are classified as repeat events.

The share of repeat purchases is computed over unique sales events rather than total quantities purchased, as a small number of high-quantity transactions would otherwise distort the analysis.

Household-Level Patterns. Repeat purchases account for a substantial share of overall activity. On average, 48.04% of items purchased by a household are products that household has bought before, with a median of 49.85% and a standard deviation of 19.23 percentage points. The distribution spans the full range from households with no repeat purchases to those in which nearly all purchases are repeats (90.77%), reflecting considerable heterogeneity in shopping behavior.

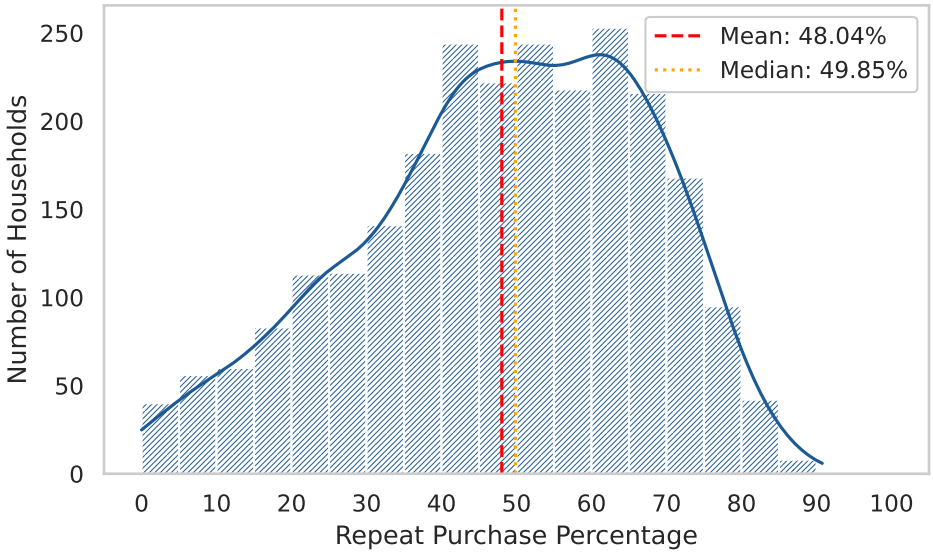

Figure 3.7: Distribution of household-level repeat purchase rates.

The majority of households fall in the 40%-70% range, while a notable share falls below, indicating more exploratory purchasing behavior. Very few households exceed an 80% repeat rate. A moderate positive correlation ($r = 0.48$) between the number of shopping occasions and the repeat purchase rate suggests that more frequent shoppers tend to develop stronger product preferences over time.

Product-Level Patterns. To assess which items drive repeat activity, products are classified as frequently purchased if a household bought them five or more times during the observation period. On average, 34.33% of the quantity purchased by a household consists of such items (SD = 17.51%, median = 33.44%), confirming that a substantial share of purchase volume is concentrated in a relatively small set of high-frequency products. Figure 3.8 shows the distribution across households.

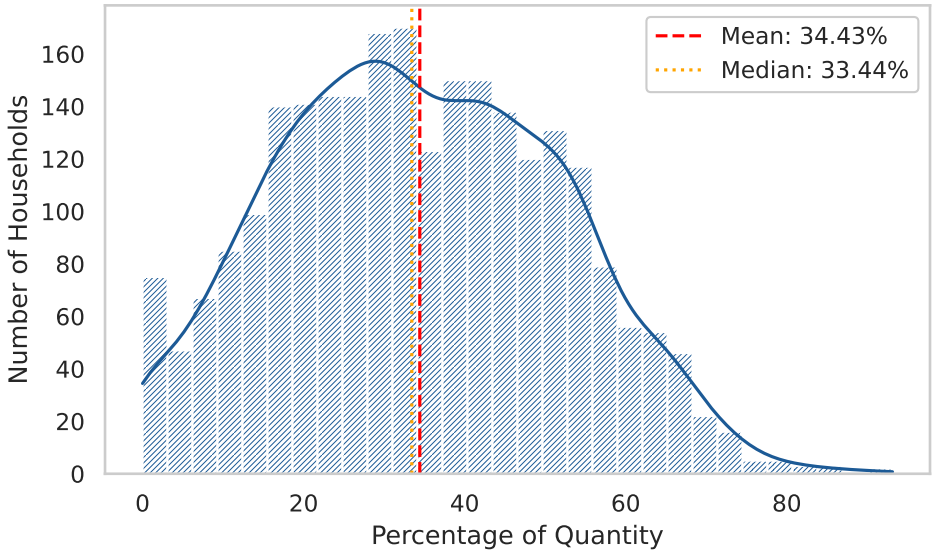

Figure 3.8: Distribution of the share of total purchased quantity attributable to frequently purchased items, across households.

3.4.4 Implications for Modeling

Three patterns emerge from this analysis that are directly relevant to the modeling approach. First, repeat purchases constitute approximately half of all transactions, motivating a recommendation strategy that gives explicit weight to repurchase likelihood. Second, the pronounced heterogeneity across households, in both repurchase rates and shopping frequency, indicates that personalized modeling is preferable to population-level approaches. Third, the concentration of purchase volume in a small set of high-frequency products suggests that a model focusing on this subset can achieve strong coverage without scoring the full catalog. These observations underpin the design choices made in the two recommendation models described in the preceding sections.

3.5 Business Value Score

3.5.1 Conceptual Foundation

Incorporating business objectives into a recommender system requires a metric that quantifies the retailer's economic interest in recommending each product. This metric must reflect the operational realities of grocery retail while remaining computationally tractable and compatible with the customer utility signal produced by the recommendation models.

To this end, we construct a business value score (BVS) that aggregates five product-level dimensions, each grounded in retail management literature and aligned with the low-margin, high-turnover economics of grocery retail. Within our multi-objective framework, the BVS serves

as the operationalization of business utility and is designed to be directly comparable to customer utility, expressed as the predicted purchase probability, to enable systematic trade-off analysis.

3.5.2 Dimensions of Business Value

The BVS comprises five dimensions, each targeting a specific operational concern.

Contribution Margin (CM). Contribution margin is defined as the difference between selling price and variable costs, expressed as a percentage of the selling price, and captures the direct profitability of each sale (Ailawadi & Farris, 2017). Given that net margins in grocery retail typically range from 1% to 3% (Trejo-Pech & White, 2024), even modest product-level margin differences have a meaningful impact on overall profitability. Margin heterogeneity is substantial across categories.

Inventory Pressure (IP). Inventory pressure quantifies the cost of holding excess stock, a concern that is particularly acute in grocery retail due to working capital constraints and limited shelf life. Products with high inventory-to-sales ratios tie up capital that could be deployed elsewhere, providing a direct economic rationale for giving such items higher priority in recommendations.

Temporal Urgency (TU). Temporal urgency reflects the economic pressure to sell perishable products as they approach expiration. Fresh produce, dairy, meat, and bakery items collectively account for approximately 20-25% of grocery sales but are responsible for 30-40% of food waste (Todd & Faour-Klingbeil, 2024; Lee, Lan, Jambrak et al., 2024). Expired inventory represents an average revenue loss of 1.5-2.5%, with perishable departments reaching 5-15% (Tabane, Phume & Retief, 2024).

Cross-Selling Potential (CSP). Cross-selling potential measures a product's tendency to stimulate complementary purchases, drawing on market basket analysis literature (Brijs et al., 1999) and the basket-oriented structure of grocery shopping (Li et al., 2009). Certain products function as anchor items that reliably co-occur with other categories. Bread purchases frequently accompany deli meat, cheese, and condiments, while pasta correlates with sauce, parmesan, and salad ingredients (Kutuzova & Melnik, 2018). Prioritizing such items in recommendations may increase overall basket value.

Strategic Priority (SP). Strategic priority captures managerial objectives that extend beyond immediate transaction economics, including promotional campaigns, supplier agreements, new product introductions, and category management targets. Retailers commonly commit to minimum shelf space or promotional support for specific suppliers, and private label penetration targets typically fall in the 15-20% range (Park et al., 2018). This dimension allows the

system to accommodate such imperatives without requiring structural changes to the recommendation pipeline.

3.5.3 Data Generation Protocol

Rationale for Simulation. The Dunnhumby dataset provides rich transactional detail but does not include the operational metrics required to compute the BVS directly: product costs, inventory levels, expiration dates, and supplier relationships are absent, a limitation common to public grocery datasets due to proprietary concerns (Ariannezhad et al., 2022).

We therefore generate synthetic business metrics by sampling from distributional ranges. This approach has precedent in recommender systems research, where business metrics are simulated to evaluate multi-stakeholder algorithms (Sürer et al., 2018; Jannach & Adomavicius, 2017). The goal of this study is not to produce accurate profit predictions for a specific retailer, but to investigate the structural trade-off between customer utility and business utility under realistic distributional assumptions. The simulation enables controlled experimentation on this trade-off while keeping the generated values consistent with empirically documented ranges.

The parameter ranges assigned to each BVS dimension are calibrated from the retail management and operations literature. Contribution margin ranges draw on the category-level margin heterogeneity documented in the grocery retail literature (Ailawadi & Farris, 2017; Trejo-Pech & White, 2024). Inventory pressure ranges are grounded in the high-turnover economics characteristic of grocery retail, where stock levels vary systematically with category velocity (Ailawadi & Farris, 2017; Tabane et al., 2024). Temporal urgency ranges follow the wide spread of shelf lives observed across grocery categories, from highly perishable fresh produce and dairy to shelf-stable packaged goods (Amorim et al., 2013; Lau et al., 2016; Herbon et al., 2014; Todd & Faour-Klingbeil, 2024). Cross-selling potential is scaled to approximate the distribution of co-occurrence strengths typically observed in grocery market basket studies, where strong product affinities cluster in a small fraction of the catalog (Brijs et al., 1999; Li et al., 2009; Kutuzova & Melnik, 2018). The strategic priority reflects the empirical observation that only a small fraction of a grocery catalog receives active strategic support at any given time, whether through promotional campaigns, supplier agreements, or category management targets (Dhar et al., 2001; Krafft & Mantrala, 2010).

Simulation Procedure. The simulation uses the `COMMODITY_DESC` attribute as its informational basis, which assigns each product a top-level category label such as `BREAD`, `COFFEE`, or `SOUP`. The procedure follows five stages.

Stage 1: Category Classification. For each BVS dimension, products are assigned to predefined clusters based on their commodity description. Clustering captures both systematic differences between categories and variation within them. For the contribution margin dimension, for instance, products are grouped into three clusters (low, medium, and high margin) according to their commodity type.

Stage 2: Parameter Range Assignment. Each dimension-cluster combination is assigned a parameter range. These ranges define the sampling from which individual product values are subsequently drawn, ensuring that generated metrics remain consistent with the operational characteristics of each cluster.

Stage 3: Value Generation. For each product and each dimension, a value is drawn from a normal distribution fitted to the assigned parameter range. This introduces within-category variation while preserving the distributional structure implied by the cluster assignment.

Stage 4: Normalization. Raw values are min-max normalized to the interval $[0, 20]$:

$$v'_{i,d} = \frac{v_{i,d} - \min_d}{\max_d - \min_d} \times 20 \tag{3.7}$$

Normalization ensures comparability across dimensions and prevents those measured in larger units from dominating the composite score.

Stage 5: Weighted Aggregation. The business value score for product i is computed as a weighted linear combination of its normalized dimension values:

$$b_i = \sum_{d=1}^{5} w_d \cdot v'_{i,d} \tag{3.8}$$

where $w_d \geq 0$ and $\sum_{d=1}^{5} w_d = 1$.

3.5.4 Weight Selection

Rather than fixing a single weighting scheme, the BVS is designed to be configured by the retailer according to current operational priorities. This preserves control and allows the system to be adapted as business priorities change. Four representative configurations are defined:

Profit-Focused: ($w_{\text{CM}} = 0.40, w_{\text{IP}} = 0.20, w_{\text{TU}} = 0.20, w_{\text{CSP}} = 0.10, w_{\text{SP}} = 0.10$). Emphasizes immediate margin contribution; suitable for financially constrained retailers.

Waste-Reduction-Focused: ($w_{\text{CM}} = 0.15, w_{\text{IP}} = 0.25, w_{\text{TU}} = 0.35, w_{\text{CSP}} = 0.10, w_{\text{SP}} = 0.15$). Prioritizes inventory turnover and perishability management; aligned with sustainability objectives.

Growth-Focused: ($w_{\text{CM}} = 0.20, w_{\text{IP}} = 0.10, w_{\text{TU}} = 0.10, w_{\text{CSP}} = 0.35, w_{\text{SP}} = 0.25$). Prioritizes basket expansion and strategic product support; suited to market share growth objectives.

Balanced: ($w_{\text{CM}} = 0.20, w_{\text{IP}} = 0.20, w_{\text{TU}} = 0.20, w_{\text{CSP}} = 0.20, w_{\text{SP}} = 0.20$). Equal weighting across all dimensions; used as the neutral baseline in subsequent experiments.

In the experiments that follow, we adopt the balanced weighting scheme as the default configuration, allowing the evaluation to focus on the structural properties of the business-customer trade-off rather than the sensitivity of results to a particular operative priority.

Symbol	Description
TIFU-KNN	
$\mathbf{v}_t$	Binary basket vector at time step t
n	Total number of historical baskets for a user
m	Number of temporal groups
g	Number of baskets per group, $g = \lceil n/m \rceil$
r_b	Within-group temporal decay ratio, $r_b \in [0, 1]$
r_g	Across-group temporal decay ratio, $r_g \in [0, 1]$
$\mathbf{u}_{\text{pers}}$	Personal user representation vector (repeated purchase component)
$\mathbf{u}_{\text{collab}}$	Collaborative user representation vector (neighbor average)
α	Fusion parameter, $\alpha \in [0, 1]$
β	Business blend parameter for TIFU-KNN adjustment, $\beta \in [0, 1]$
ReCANet	
$\vec{h}$	History vector encoding temporal gaps between repeat purchases
W	Window size hyperparameter (length of $\vec{h}$)
L	Number of most recent baskets used for training
$\mathcal{T}$	Set of all user-item training pairs
$\mathcal{L}$	Loss function (binary cross-entropy or business-aware variant)
Business Value Score	
$v_{i,d}$, $v'_{i,d}$	Raw and normalized value of item i on BVS dimension d
w_d	Weight for BVS dimension d, with $\sum_d w_d = 1$
MIP Optimization	
x_i	Binary decision variable: 1 if item i is recommended, 0 otherwise
f_{cust}, f_{bus}	Customer and business utility objective functions
λ	Weighting parameter for weighted sum scalarization, $\lambda \in [0, 1]$
ε	Minimum business value threshold (ε-constraint method)
Evaluation	
B_u^{test}	Set of items in user u's held-out test basket
π_j	Item at rank position j in the recommendation list

Table 3.2: Unified notation: model-specific and method-specific symbols.

4 Business-Awareness Integration Methods

4.1 Multiplicative Scoring

The multiplicative scoring approach is a straightforward baseline for incorporating business objectives into a recommender system. It operates as a post-processing reranking method on the output of any standard recommender system.

A conventional recommender system produces a ranked list of products ordered by their predicted purchase probability. The product most likely to be purchased is ranked first. In the multiplicative scoring approach, this purchase probability is multiplied by the business score of the respective product. The resulting fused score combines the predicted customer preference with the economic value of selling the product. The recommendation list is then reranked according to this fused score.

4.1.1 Problem Formulation

Let $\hat{p}_{u,i}$ denote the prediction score of a standard recommender system, representing the predicted purchase probability or preference score for user u and product i. Each product i is further associated with a business score b_i as defined previously.

The multiplicative scoring method computes a fused score $S_{u,i}$ as follows:

$$S_{u,i} = \hat{p}_{u,i} \cdot b_i \tag{4.1}$$

The set of top-k recommendations for user u is then determined by selecting the products with the highest fused scores:

$$\hat{R}_u^k = \text{top-}k\{i : S_{u,i}\}, \quad i \in I \tag{4.2}$$

where I denotes the set of all available products.

4.1.2 Theoretical Properties

Monotonicity. The multiplicative fusion preserves monotonic relationships. If item i_1 dominates item i_2 on both dimensions ($\hat{p}_{u,i_1} \geq \hat{p}_{u,i_2}$ and $b_{i_1} \geq b_{i_2}$), then $S_{u,i_1} \geq S_{u,i_2}$. This ensures that Pareto-dominant items are never displaced by Pareto-dominated alternatives.

Trade-off characteristics. The multiplicative function exhibits substitutability between objectives. An item can achieve a high fused score either by excelling on one dimension or by achieving moderate scores on both. Constant fused scores define hyperbolic indifference curves in the $(\hat{p}_{u,i}, b_i)$ space:

$$b_i = S_{u,i}/\hat{p}_{u,i} \tag{4.3}$$

This implies increasing marginal rates of substitution, since as the business value decreases, increasingly large improvements in relevance are needed to maintain the same fused score.

Implicit weighting. The multiplicative fusion implicitly weights objectives based on the scale and distribution of scores. If business scores exhibit higher variability than relevance scores, business value effectively receives greater weight in determining the ranking. This lack of explicit control over the utility trade-off is a limitation compared to parameterized methods.

By multiplying the two components, products that are both relevant to the customer and valuable to the business receive the highest combined scores. Products that score poorly on either dimension are naturally deprioritized. Since the method operates purely as a post-processing layer, it requires no modifications to the underlying recommender system architecture.

4.1.3 Practical Implementation

The deployment of the multiplicative scoring method is straightforward and requires negligible computational overhead. Since it operates as a model-agnostic post-processing step, it can be applied to the output of any baseline recommender system.

As detailed in Algorithm 1, the procedure requires an element-wise multiplication of the score arrays followed by a standard sort.

Algorithm 1: Multiplicative Scoring Reranking

Input : Set of all available items I, Business scores B, list size k
Output : Recommended list of k items

for *each item* $i \in I$ **do**
  $S_{u,i} \leftarrow \hat{p}_{u,i} \cdot b_i$;
end
return Top k items from I sorted by $S_{u,i}$;

4.1.4 Advantages and Limitations

The method requires no modification to existing recommender systems and functions as a post-processing layer with minimal implementation effort. It is entirely model-agnostic, working with any base recommender that produces numerical scores, and adds negligible computational overhead. The fused score also has an intuitive interpretation as the expected business value of a recommendation, which aligns with decision-theoretic principles of expected utility maximization.

At the same time, the method has several notable limitations. Most importantly, the multiplicative fusion provides no explicit control over the trade-off between customer relevance and

business value. The multiplicative operation creates hard exclusions. Any item with a relevance or business score of zero receives a fused score of zero regardless of its value on the other dimension. Moreover, items with very low business value cannot be recommended even if they are highly relevant to the customer.

Despite these limitations, multiplicative fusion is a useful baseline due to its simplicity and its value as a reference point for evaluating more sophisticated approaches. It represents the minimal-complexity approach to business-aware recommendation, establishing a performance floor against which the investment in more complex methods can be assessed.

4.2 Mixed Integer Programming

Following the multiplicative baseline, the second method frames the integration of business awareness as a multi-objective optimization problem solved via Mixed Integer Programming (MIP). Like the previous approach, this method operates as a post-processing reranking step. However, rather than relying on a heuristic fusion of scores, it explicitly models the trade-off between customer relevance and business value.

4.2.1 Problem Formulation

We define the recommendation task as a selection problem in which a subset of items is to be chosen for recommendation to a user u from the set of all available items I. To this end, we introduce a binary decision variable x_i for each item $i \in I$:

$$x_i = \begin{cases} 1, & \text{if item } i \text{ is recommended} \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

The problem is characterized by two objective functions that are to be maximized simultaneously.

Objective 1: Customer Utility (f_{cust}). We assume that recommending products with high predicted purchase probability maximizes the utility for the customer. The first objective therefore maximizes the sum of the purchase probabilities ($\hat{p}_{u,i}$) of the selected items:

$$\max f_{\text{cust}}(\mathbf{x}) = \sum_{i \in I} \hat{p}_{u,i} \cdot x_i \tag{4.5}$$

Objective 2: Business Utility (f_{bus}). The second objective captures the economic interest of the company. It maximizes the total business value of the recommendation list by summing the business scores (b_i) of the selected items:

$$\max f_{\text{bus}}(\mathbf{x}) = \sum_{i \in I} b_i \cdot x_i \tag{4.6}$$

Constraints. The primary constraint in this recommendation setting is the size of the recommendation list, formulated as a cardinality constraint. For a top-k list (e.g., $k = 5$), the constraint reads:

$$\sum_{i \in I} x_i = k \tag{4.7}$$

4.2.2 Resolution Strategy: Scalarization

Since both objective functions are linear, we employ scalarization methods to convert the multi-objective problem into a single-objective formulation that can be solved with standard solvers. To explore the trade-offs and compute the Pareto frontier, the set of optimal solutions where one objective cannot be improved without degrading the other, we consider two methods.

Weighted Sum Method. In this approach, the two objective functions are combined into a single utility function using a weighting parameter $\lambda \in [0, 1]$. This parameter controls the relative importance of customer utility versus business utility.

$$\begin{aligned} \max \quad & Z(\mathbf{x}) = \lambda \cdot f_{\text{cust}}(\mathbf{x}) + (1 - \lambda) \cdot f_{\text{bus}}(\mathbf{x}) \\ \text{s.t.} \quad & \sum_i x_i = k \\ & x_i \in \{0, 1\} \quad \forall\, i \in I \end{aligned} \tag{4.8}$$

By varying λ, different recommendation lists can be generated that range from purely customer-centric to purely business-centric.

Epsilon-Constraint Method. As a complementary approach, we employ the ε-constraint method. Here, one objective function is optimized while the other is imposed as a constraint with a lower bound ε:

$$\begin{aligned} \max \quad & f_{\text{cust}}(\mathbf{x}) \\ \text{s.t.} \quad & f_{\text{bus}}(\mathbf{x}) \geq \varepsilon \\ & \sum_{i \in I} x_i = k \end{aligned} \tag{4.9}$$

By systematically varying ε, the solution space is mapped out, ensuring a guaranteed level of business value while maximizing customer relevance.

4.2.3 Theoretical Properties

Optimality guarantees. Unlike heuristic reranking methods, the MIP formulation provides provable optimality. For a given weight λ, the solution is guaranteed to be the best possible recommendation set balancing the two objectives:

$$\nexists\, S \subset I \text{ with } |S| = k \text{ such that } Z(S) > Z(\hat{R}_u^k) \tag{4.10}$$

This property enables confident interpretation of trade-off curves and Pareto frontiers.

Pareto efficiency. All solutions generated by varying $\lambda \in [0, 1]$ lie on the Pareto frontier, meaning no solution can improve one objective without worsening the other. This ensures that the identified trade-offs are truly optimal rather than suboptimal compromises.

Computational complexity. Because the only constraint is a uniform cardinality limit, the constraint matrix is totally unimodular. The LP relaxation therefore yields exact integer solutions, and the base formulation is solvable in polynomial time.

4.2.4 Practical Implementation

Translating the MIP formulations into a practical system requires addressing computational latency. To this end, the reranking logic operates on a truncated candidate set containing the top C recommendations rather than the full item set.

Data normalization. Since predicted purchase probabilities ($\hat{p}_{u,i}$) and business scores (b_i) operate on different scales, directly combining them leads to instability. Before applying either reranking strategy, we therefore apply min-max normalization to bound both sets of values to the interval $[0, 1]$.

Weighted Sum Algorithm. While the weighted sum formulation is theoretically a Mixed Integer Program, implementing a dedicated MIP solver for this specific problem is unnecessary. Because the objective function is purely linear and the only constraint is a uniform cardinality limit ($\sum x_i = k$), the problem satisfies the properties of a totally unimodular matrix. The exact optimal solution can therefore be found by sorting. As detailed in Algorithm 2, calculating the combined utility score for each item and selecting the top-k yields the optimal solution.

Algorithm 2: Weighted Sum Reranking

Input : Candidate set I_C, Business scores B, list size k, weight λ, candidate size C
Output : Recommended list of k items

$I_{\text{top}} \leftarrow$ Top C items from I_C sorted by purchase probability;
$P_{\text{norm}} \leftarrow \text{MinMaxNormalize}(I_{\text{top}}.\text{probabilities})$;
$B_{\text{norm}} \leftarrow \text{MinMaxNormalize}(I_{\text{top}}.\text{business\_scores})$;
for *each item* $i \in I_{top}$ **do**
  $\text{Score}_i \leftarrow \lambda \cdot P_{\text{norm}}[i] + (1 - \lambda) \cdot B_{\text{norm}}[i]$;
end
return Top k items from I_{top} sorted by Score;

Epsilon-Constraint Algorithm. Unlike the weighted sum method, the ε-constraint method introduces a non-uniform knapsack-like constraint ($\sum b_i \cdot x_i \geq \varepsilon \cdot k$). A greedy sorting approach cannot guarantee that this lower bound on business value is satisfied while maximizing customer utility. This method therefore requires a dedicated MIP solver.

As shown in Algorithm 3, the absolute ε threshold is calculated dynamically based on the requested percentile of the candidate set's business scores. If the top candidates are inherently

low in business value such that the ε constraint becomes infeasible, the system defaults to the baseline probability-ranked list.

Algorithm 3: Epsilon-Constraint Reranking

Input : Candidate set I_C, Business scores B, list size k, percentile P_ε, candidate size C
Output : Recommended list of k items

$I_{\text{top}} \leftarrow$ Top C items from I_C sorted by purchase probability;
Impute missing values in B using $\text{median}(B)$;
$\varepsilon \leftarrow \text{CalculatePercentile}(I_{\text{top}}.\text{business\_scores},\ P_\varepsilon)$;
Initialize exact MIP Solver (Maximize);
Define binary decision variables $x_i \in \{0, 1\}$ for each item $i \in I_{\text{top}}$;
Objective: Maximize $\sum_i \left(I_{\text{top}}.\text{probabilities}[i] \cdot x_i\right)$;
Constraint 1: $\sum_i x_i = k$;
Constraint 2: $\sum_i \left(I_{\text{top}}.\text{business\_scores}[i] \cdot x_i\right) \geq \varepsilon \cdot k$;
Solve MIP;
if *Solver Status is Optimal* **then**
    return items where $x_i = 1$;
else
    return Top k items from I_{top} sorted purely by purchase probability;
end

4.2.5 Advantages and Limitations

The MIP formulation offers several advantages over heuristic approaches. For a given λ or ε, the solution is provably optimal, it is the best possible trade-off between customer and business utility. This optimality enables the generation of complete Pareto frontiers through systematic variation of the parameters, providing a comprehensive characterization of the trade-off space. The explicit parameters λ and ε offer direct, interpretable control over the balance between objectives, and the framework can be extended with additional constraints such as category diversity requirements.

These advantages come with practical costs. The primary limitation is computational expense. While the problem sizes considered here ($|I| \approx 10^4, k = 10$) can be solved in seconds, the NP-hard nature of general MIP formulations limits scalability to larger catalogs or real-time optimization. Moreover, the method requires selecting appropriate values for λ or ε, which introduces a meta-optimization problem.

The MIP approach is therefore best suited for scenarios where optimality guarantees are valuable, such as strategic planning or benchmark establishment, and where computational budgets permit the optimization overhead.

4.3 TIFU-KNN Model Adjustment

The standard TIFU-KNN algorithm constructs a user representation vector that encodes purchase history through temporal weighting. Each item occurrence is weighted according to the recency of the basket in which it was observed. Business-relevant properties of items play no role in this representation. Consequently, a low-business-value product that was purchased repeatedly will dominate a user's vector and, by extension, the neighbor selection and prediction process, regardless of its commercial value to the retailer.

We introduce a *Business-Aware Decay* extension to TIFU-KNN that addresses this limitation by integrating business value directly into the user representation layer rather than applying it as a post-processing step. The core idea is to replace the purely temporal item weight with a blended weight that combines temporal decay with the item-level business score. This modification reshapes the internal representation of every user such that high-value items are systematically amplified and low-value ones are softly suppressed, while the full computational structure and interpretability of the original algorithm are preserved.

4.3.1 Business Value Integration Strategy

The integration operates at the level of the item weight assigned within each basket during the construction of the user history vector. In standard TIFU-KNN, the contribution of item i in basket t to the user vector is determined solely by a temporal decay factor that reflects the distance of basket t from the most recent observation. The business-aware extension introduces a blending multiplier that modulates this contribution as a continuous function of the item's business score.

Each item i in the product catalog is assigned a normalized business weight $w_i \in [0, 1]$, derived from its business value score b_i through min-max normalization over all business scores. The blending multiplier for item i is then defined as:

$$\text{blend}(i) = (1 - \beta) + \beta \cdot w_i \tag{4.11}$$

where $\beta \in [0, 1]$ is the business blend parameter controlling the strength of business value integration. The multiplier is bounded between $(1 - \beta)$ and 1.0 for all items, ensuring that no item is completely suppressed unless $\beta = 1$ and $w_i = 0$ simultaneously.

4.3.2 Problem Formulation

The modification is applied within the user vector computation. For each basket in the user's history, the standard temporal decay value is computed as in the original algorithm. The contribution of item i to the basket vector is then given by:

$$\text{his\_vec}[i] = \text{decay}(t) \cdot \text{blend}(i) = r_b^{(n-t-1)} \cdot [(1 - \beta) + \beta \cdot w_i] \tag{4.12}$$

where $r_b \in [0, 1]$ is the within-group decay rate, n is the total number of historical baskets, and t indexes baskets from oldest ($t = 0$) to most recent ($t = n - 1$). All subsequent steps of the al-

gorithm, including group-level temporal weighting, KDTree construction for neighbor retrieval, and the α-weighted combination of personal and collaborative components, proceed identically to standard TIFU-KNN. These steps operate on the business-inflected vectors throughout.

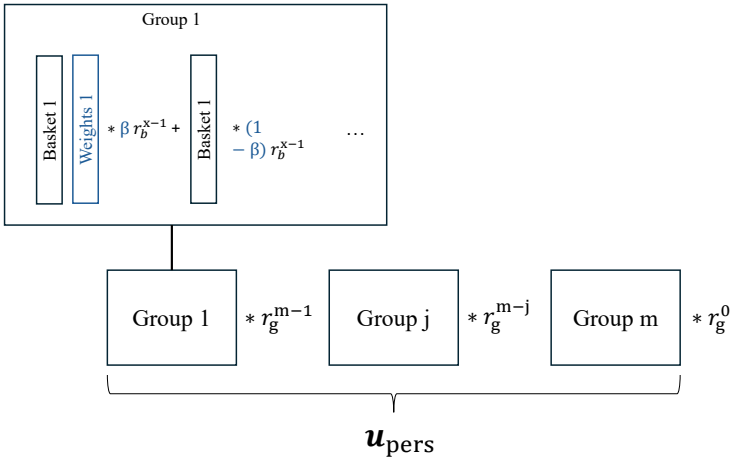

Figure 4.1: Each basket is additionally weighted by a vector of item-level business scores, controlled by blend parameter β to create the business-aware user representation.

Because the KDTree is built from business-aware user vectors, neighbor selection is also influenced by the modification. Users who purchased similar high-value items are drawn closer together in the representation space, while users whose histories are dominated by low-value items become relatively more distant from high-value buyers. Unlike an approach that would filter only the collaborative component while leaving the personal component unmodified, which would create a representation mismatch between the two halves of the prediction, the business-aware decay applies consistently across both components. Both personal history and collaborative signals are derived from the same representation, ensuring internal coherence.

4.3.3 Theoretical Properties

The behavior of the blend parameter β can be characterized across its full range. At $\beta = 0$, $\text{blend}(i) = 1.0$ for every item regardless of business score, and the algorithm reduces exactly to standard TIFU-KNN. This boundary case serves as a built-in verification. Experiments at $\beta = 0$ should reproduce baseline performance precisely.

At intermediate values, the multiplier introduces a graduated re-weighting. For $\beta = 0.3$, a zero-value item ($w_i = 0$) receives a multiplier of 0.70, a median item ($w_i = 0.5$) receives 0.85,

and a maximum-value item ($w_i = 1.0$) retains the full multiplier of 1.00. The spread between the lowest and highest multipliers is 0.30, meaning business value introduces a 30% differential in item weight relative to what pure temporal decay would assign.

At $\beta = 1.0$, the multiplier equals w_i directly, so the effective item weight becomes the product of temporal decay and normalized business score. Items with $w_i = 0$ contribute nothing to the user representation regardless of recency or frequency of purchase. This is the most aggressive business-aware configuration, analogous in effect to a hard threshold applied at the representation level.

A key property of this formulation is that the business influence propagates through the full model rather than being confined to a single component. Because every user vector, both those stored in the KDTree and those computed at inference time, is built from the same blended weights, the personal and collaborative components of the final prediction are structurally consistent. The parameter β therefore exerts influence through three simultaneous channels: the target user's personal score vector, the set of neighbors selected as similar users, and the neighbor average used as the collaborative prediction.

4.3.4 Balancing Customer and Business Utility

The parameter β provides a single, interpretable control for the customer-business trade-off. At low values, the model behaves nearly identically to standard TIFU-KNN, with business value exerting only a mild influence on item weights. As β increases, the model progressively shifts its internal representation toward a business-value-weighted view of purchase history, at the cost of diverging from the raw preference signal that drives recommendation accuracy.

The approach is non-exclusionary. Even at high β values, low-value items remain present in user vectors with reduced but non-zero weight (unless $\beta = 1$ and $w_i = 0$ exactly). A highly relevant low-value item that a user has purchased repeatedly and recently can therefore still achieve a competitive score, because its temporal signal is strong enough to compensate for the reduced business weight. The model does not impose an absolute veto on any item but adjusts the effective evidence for each item continuously as a function of both recency and business value.

This soft, continuous trade-off mechanism distinguishes the business-aware decay approach from post-processing reranking methods. The business signal is embedded in the representation itself, making it structurally inseparable from the temporal signal rather than being layered on top of a pre-computed relevance ranking.

4.4 ReCANet Model Adjustment

The ReCANet algorithm generates recommendations through a deep learning architecture that integrates static and sequential user features into a unified prediction model. Users and items are mapped to dense embedding vectors to capture latent characteristics, while the user's sequential purchase history is processed through a Long Short-Term Memory (LSTM) layer to capture temporal dependencies and evolving preferences. These static and sequential representations are concatenated and passed through a Multi-Layer Perceptron (MLP), where the final

layer applies a sigmoid activation function to output a probability score representing the likelihood of a specific user interacting with a given target item.

In the standard implementation, the model is optimized to minimize prediction error uniformly across all items, treating every potential interaction with equal importance regardless of its economic implications.

4.4.1 Business-Aware Extension

We introduce a modification to the ReCANet training process that incorporates business objectives directly into the neural network's optimization. Rather than adjusting predictions after training, this extension modifies the learning phase itself by exploiting the loss function as the steering mechanism of gradient-based optimization. The objective is to create a cost-sensitive learning environment in which the model prioritizes learning patterns associated with high-value items while maintaining its capacity to identify strong customer preferences for standard items.

4.4.2 Business Value Integration Strategy

The business-aware extension operates by introducing sample-level weighting to the objective function, altering how the model calculates error gradients during backpropagation. This shifts the model's optimization target from uniform prediction accuracy across all items to value-weighted accuracy.

Each item i in the catalog is assigned a business weight w_i derived from its business value score b_i. To ensure comparability across items and numerical stability during training, raw business value scores are first min-max normalized to the unit interval:

$$\tilde{b}_i = \frac{b_i - \min_{j \in \mathcal{I}} b_j}{\max_{j \in \mathcal{I}} b_j - \min_{j \in \mathcal{I}} b_j}, \quad \tilde{b}_i \in [0, 1] \tag{4.13}$$

The business weight is then defined as an exponential function of the normalized score:

$$w_i = \exp\left(\lambda \cdot \tilde{b}_i\right) \tag{4.14}$$

where $\lambda \in \mathbb{R}$ is the weighting parameter that controls the strength of business-aware integration. The exponential form is chosen because it produces a multiplicative effect on gradients, providing strong differentiation between high- and low-value items at moderate λ values without requiring extreme parameter settings.

At $\lambda = 0$, all items receive a uniform weight of $w_i = 1$, and the loss function reduces exactly to the standard binary cross-entropy of the unmodified ReCANet model. At positive values of λ, high-BVS items receive disproportionately larger weights, incentivizing the model to allocate greater representational capacity to their correct prediction.

4.4.3 Problem Formulation

The modification replaces the standard binary cross-entropy loss with a weighted variant. During training, the loss contribution of each sample is scaled by the business weight of the target item.

Weighted Objective Function. We introduce item-specific weighting to the loss calculation. For a batch of N samples, the weighted loss function is defined as:

$$\mathcal{L}_{\text{BA}} = -\frac{1}{N} \sum_{(u,i)\in\mathcal{T}_N} w_i \cdot [y_{u,i} \log(\hat{p}_{u,i}) + (1 - y_{u,i}) \log(1 - \hat{p}_{u,i})] \tag{4.15}$$

Gradient Scaling Mechanism. The weighted loss function directly impacts the gradient descent process. The gradient of the loss with respect to the model parameters θ becomes:

$$\nabla_\theta \mathcal{L}_{\text{BA}} \approx w_i \cdot (\hat{p}_{u,i} - y_{u,i}) \cdot \nabla_\theta \sigma(z_i) \tag{4.16}$$

The magnitude of each weight update is thus scaled by w_i. Errors on high-value items result in larger gradients, causing the optimizer to adjust the network parameters more substantially to correct them.

4.4.4 Theoretical Properties

Impact on Learning Dynamics. The business-aware modification affects the model's convergence behavior in a predictable manner. The model is incentivized to distribute its finite capacity such that the weighted cumulative error is minimized:

$$\mathbb{E}[\mathcal{L}_{\text{BA}}] \approx \sum_{i\in I} p(i) \cdot w_i \cdot \text{Error}(i) \tag{4.17}$$

This creates a risk-averse learning dynamic with respect to high-value items. Failing to predict a relevant high-value item (a false negative where w_i is large) incurs a substantially higher penalty than failing to predict a low-value item. Conversely, the model is permitted to be less confident about low-value items unless the signal from the user history is unambiguous.

Trade-off Control. The method provides an implicit mechanism for controlling the customer-business trade-off through the distribution of weights. Unlike the explicit threshold behavior of the TIFU-KNN adjustment, this weighting allows for a continuous trade-off. A strong customer signal can still overcome a low business weight, preserving relevance for low-margin items.

4.4.5 Balancing Customer and Business Utility

This approach achieves a balance between customer utility and business utility through probabilistic prioritization rather than exclusionary filtering. The model retains the ability to recom-

mend any item in the catalog, ensuring that the full range of user preferences remains accessible.

The resulting recommendation behavior exhibits the following properties. Items with high business value are learned with higher confidence, making them more robust to noise and more likely to appear in top-k lists when user intent is ambiguous. Items with lower business value are not discarded but require a stronger signal from the user's sequential history to achieve a probability score sufficient for inclusion in the top-k list. The mechanism effectively creates a relevance threshold that scales with business value. Low-value items must be highly relevant to be recommended, while high-value items can be recommended at slightly lower, but still positive, relevance levels.

4.5 Evaluation Framework

To systematically assess the performance of the baseline and business-aware recommender systems, we employ a dual evaluation framework that separately quantifies customer utility and business utility. This separation reflects the multi-objective nature of the problem. A system that improves business value at the cost of recommendation quality is not necessarily preferable to one that achieves a more modest but balanced improvement on both dimensions. The evaluation framework is therefore designed to measure each objective in isolation and to support the trade-off analysis that is central to the research questions.

4.5.1 Evaluation Protocol

The dataset is split at the user level using a leave-last-basket-out strategy. For each user $u \in U$, all baskets except the final observed basket are used for training, and the final basket constitutes the ground truth against which recommendations are evaluated. This protocol is standard in next-basket recommendation research (Ariannezhad et al., 2022) and is well suited to the grocery retail context, where the objective is to predict what a customer will purchase on their next shopping trip given their full prior history.

Evaluation is performed in a macro-averaged fashion. All metrics are first computed per user and then averaged across all users in the test set, assigning equal weight to each user regardless of purchase frequency or history length. This ensures that the reported metrics reflect typical performance across the user population rather than being dominated by a small number of highly active households.

All experiments use a consistent subset of 1,500 users over a one-year observation window, as determined by the sliding window analysis described in Section 5.1. Results are reported for recommendation list sizes $k \in \{5, 10, 20, 50\}$ to reflect the range of practical deployment contexts discussed below.

4.5.2 Customer Utility Metrics

Customer utility is measured using two complementary ranking metrics, Recall@k and Normalized Discounted Cumulative Gain (NDCG@k), both of which are standard in the next-basket

recommendation literature (Ariannezhad et al., 2022; Jannach & Abdollahpouri, 2023). In both metrics, relevance is binary as an item is considered relevant if and only if it appears in the user's held-out last basket.

Recall@k. Recall@k measures the proportion of items in the user's actual next basket that were successfully recovered by the top-k recommendation list:

$$\text{Recall@}k = \frac{1}{|U|} \sum_{u \in U} \frac{|\hat{R}_u^k \cap B_u^{\text{test}}|}{|B_u^{\text{test}}|} \tag{4.18}$$

where $\hat{R}_u^k$ denotes the set of top-k recommended items for user u and B_u^{test} denotes the set of items in the user's held-out basket. Recall captures coverage, how much of what the customer actually wanted is included in the recommendation list. It does not account for the position at which a relevant item appears, treating all ranks within the top-k equally.

Recall naturally increases with k, as a longer recommendation list has more opportunities to include items from the ground truth basket. Absolute Recall values are therefore not directly comparable across different values of k, and results should be interpreted within each k separately. The choice of multiple k values reflects the different practical contexts in which recommendations can be deployed. A list of $k = 5$ corresponds to a compact recommendation widget on a mobile screen, while $k = 50$ corresponds to a full personalized product feed. Each setting imposes different constraints on the user experience and on the trade-off between customer and business objectives.

NDCG@k. Normalized Discounted Cumulative Gain (NDCG@k) extends Recall by incorporating the rank position of relevant items, rewarding systems that place relevant items higher in the recommendation list:

$$\text{NDCG@}k = \frac{1}{|U|} \sum_{u \in U} \frac{\text{DCG@}k_u}{\text{IDCG@}k_u} \tag{4.19}$$

where the Discounted Cumulative Gain for user u is:

$$\text{DCG@}k_u = \sum_{j=1}^{k} \frac{\mathbb{1}[\pi_j \in B_u^{\text{test}}]}{\log_2(j+1)} \tag{4.20}$$

and IDCG@k_u is the ideal DCG achieved by a perfect ranking that places all relevant items at the top of the list, used to normalize the score to $[0, 1]$. Here, π_j denotes the item at rank position j in the recommendation list for user u, and $\mathbb{1}[\cdot]$ is the indicator function. The logarithmic discount factor $\log_2(j+1)$ assigns progressively less credit to relevant items at lower positions, reflecting the observation that users are less likely to engage with recommendations further down a list.

As with Recall, NDCG values are not directly comparable across different values of k in absolute terms, since both the DCG and the IDCG change as the list length grows. Within a fixed k, however, NDCG provides a more nuanced view of recommendation quality than Recall. Two

systems with identical Recall@k can differ substantially in NDCG if one consistently places relevant items near the top of the list while the other places them near the bottom.

Together, Recall@k and NDCG@k capture complementary aspects of customer utility. Recall measures whether the relevant items are present in the recommendation list, while NDCG measures whether they are presented in an appropriate order.

4.5.3 Business Utility Metric

Business utility is measured using the aggregate business value score of the recommendation list, referred to as BizValue@k:

$$\text{BizValue@}k = \frac{1}{|U|} \sum_{u \in U} \sum_{i \in \hat{R}_u^k} b_i \tag{4.21}$$

where b_i is the business value score of item i as defined in Section 3.5, and $\hat{R}_u^k$ is the top-k recommendation list for user u. BizValue@k represents the average total commercial utility delivered by the recommendation list across all users. A higher value indicates that the system tends to recommend items with greater economic importance to the retailer, whether through higher contribution margins, inventory pressure, temporal urgency, cross-selling potential, or strategic priority.

As with the customer utility metrics, BizValue@k increases with k because longer lists contain more items and therefore accumulate more business value. Absolute values are not directly comparable across list sizes, and results should be interpreted within each k. The unmodified baseline systems serve as the primary reference point. The relevant question is not the absolute BizValue level, but how much it changes when business-aware modifications are applied.

4.5.4 Trade-off Analysis and the Pareto Frontier

Because the evaluation framework tracks two objectives simultaneously, no single number summarizes system performance. A business-aware system is only meaningfully better than the baseline if it improves BizValue without incurring an unacceptable degradation in customer utility. Assessing this requires explicit trade-off analysis.

As introduced in Section 4.2, a solution is Pareto-optimal if no other solution can improve one objective without worsening the other. The set of all Pareto-optimal solutions forms the Pareto frontier, which represents the best achievable trade-off curve between customer utility and business utility for a given recommender model and integration method. Points on the frontier are not ranked against each other in absolute terms; the choice among them is a business decision that depends on how much accuracy loss a retailer is willing to accept in exchange for a given gain in business value.

The Pareto frontier is traced empirically by varying the control parameter of each integration method across its full range. For the weighted sum method, this parameter is $\lambda \in [0, 1]$; for the epsilon-constraint method, it is ε; and for the model adjustment methods, the corresponding

parameters are β (TIFU-KNN) and w_i (ReCANet). Each parameter setting produces a single operating point in the (customer utility, business utility) space, and the collection of these points across the parameter range traces the achievable trade-off curve for that method and model combination.

5 Results and Discussion

5.1 Baseline Performance

To establish a performance baseline for the subsequent experiments, we first evaluate both recommender systems in their unmodified form. The underlying dataset comprises 2,500 users over a period of two years. To identify a suitable data configuration, we apply a sliding window approach in which both the number of users and the length of the time window are varied. This analysis indicates that a subset of 1,500 users over a one-year period yields the most stable and representative results. We therefore use this configuration throughout all experiments.

For each model, we perform a grid search optimizing for Recall and NDCG and report the best configurations below.

5.1.1 TIFU-KNN

TIFU-KNN is a non-neural collaborative filtering model based on k-nearest neighbors with time-decay weighting. Despite its simplicity, the model is computationally lightweight and achieves competitive performance. The grid search yields the following optimal hyperparameter configuration:

n_neighbors	within_decay_rate	group_decay_rate	alpha	n_groups
100	0.9	0.7	0.7	3

Table 5.1: Optimal hyperparameters for TIFU-KNN.

The high decay rates indicate that recent purchases carry substantial weight in the prediction, which is consistent with the repetitive purchasing patterns typically observed in grocery retail. With this configuration, the model achieves the performance shown in Table 5.2.

K	Recall	NDCG	BizValue
5	0.1084	0.1885	177.46
10	0.1513	0.1802	349.77
20	0.1954	0.1783	687.68
50	0.2573	0.1923	1690.62

Table 5.2: Baseline performance of TIFU-KNN.

Figure 5.1 shows the median predicted score at each ranking position across all recommendations generated by TIFU-KNN. The distribution exhibits a moderate initial decline at the top ranks, followed by an extended tail that decreases gradually across the remaining candidate pool. Scores approach but do not reach zero even at the lowest ranks.

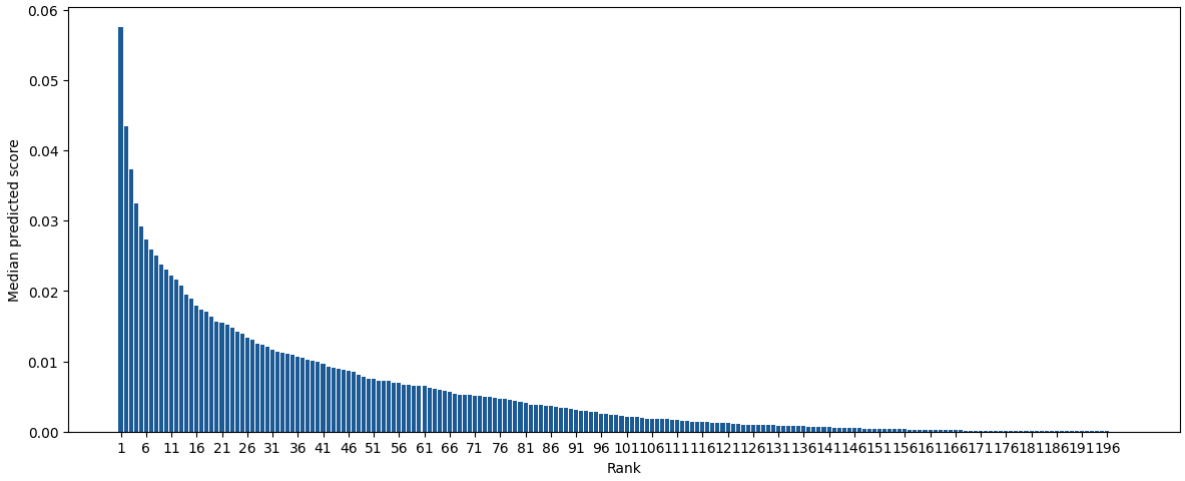

Figure 5.1: TIFU-KNN: Median predicted scores by rank.

5.1.2 ReCANet

ReCANet is a neural network-based model that captures sequential purchasing behavior through learned user and item embeddings. As noted in the original publication, embedding dimensionality has relatively little impact on performance, which allows us to keep these values small and focus the grid search on the history length. The resulting optimal configuration is given in Table 5.3.

embedding_dim	hidden_units	history_length
16	64	25

Table 5.3: Optimal hyperparameters for ReCANet.

The history length of 25 means that the recommender considers the last 25 baskets of each customer. With an average of approximately 80 baskets per user, this further underlines the higher relevance of more recent purchasing behavior. The resulting performance is reported in Table 5.4.

K	Recall	NDCG	BizValue
5	0.1215	0.1980	182.30
10	0.1711	0.1952	357.85
20	0.2150	0.1955	696.88
50	0.2773	0.2101	1641.52

Table 5.4: Baseline performance of ReCANet.

Figure 5.2 shows the median predicted score distribution across ranked candidates for ReCANet. The distribution is heavily concentrated at the top, with the highest-ranked item receiving a median score of approximately 0.31. Scores decline rapidly within the first 20 to 30 ranks and then flatten, remaining close to zero across the remainder of the candidate pool.

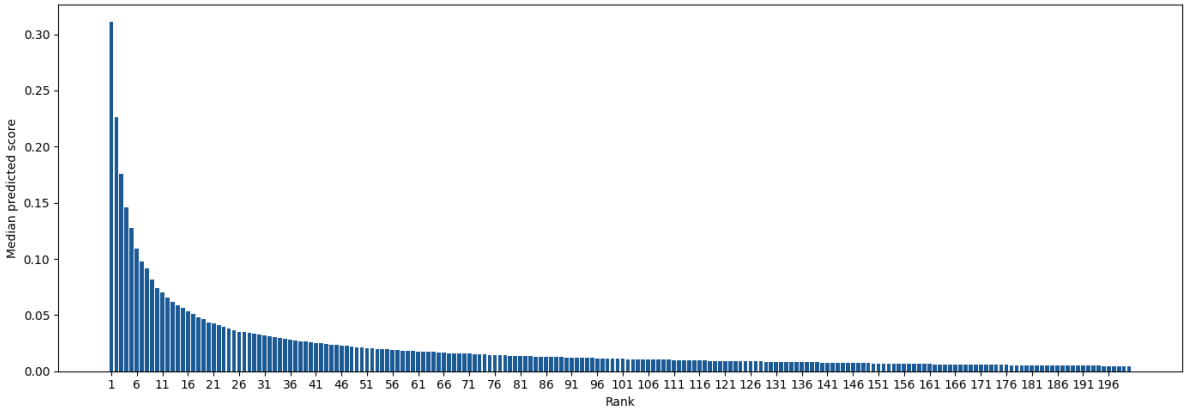

Figure 5.2: ReCANet: Median predicted scores by rank.

5.1.3 Discussion

ReCANet outperforms TIFU-KNN across all values of K on both Recall and NDCG. This is expected given the greater model capacity of the neural network. For instance, at $K = 50$, ReCANet achieves a Recall of 0.277 compared to 0.257 for TIFU-KNN and consistently higher NDCG values across all cutoffs. The business values reported here serve as a reference for the subsequent experiments. It should be noted that standard recommender models do not incorporate business value into their objective and therefore do not optimize for it.

Despite its lower predictive performance, TIFU-KNN remains a valuable point of comparison. It requires considerably fewer computational resources while remaining competitive, which makes it a relevant alternative in resource-constrained settings.

5.2 Multiplicative Scoring

The multiplicative scoring approach serves as the first business-aware modification evaluated in this work. As described in Section 4.1, it operates as a post-processing reranking step that requires no changes to the underlying recommender systems. For each user, the predicted purchase probabilities generated by the base model are multiplied element-wise with the business value scores, and the recommendation list is reordered accordingly.

5.2.1 TIFU-KNN

Applying multiplicative scoring to TIFU-KNN yields consistent improvements in BizValue across all values of K, at the cost of moderate reductions in Recall and NDCG. The results are summarized in Table 5.5 and visualized in Figure 5.3.

Table 5.5: Performance of TIFU-KNN under multiplicative scoring.

K	Rec (Orig)	Rec (Rer.)	Rec $\Delta\%$	NDCG (Orig)	NDCG (Rer.)	NDCG $\Delta\%$	Biz (Orig)	Biz (Rer.)	Biz $\Delta\%$
5	0.1084	0.1000	−7.76%	0.1885	0.1868	−0.93%	177.46	196.79	+10.89%
10	0.1513	0.1457	−3.68%	0.1802	0.1768	−1.90%	349.77	385.77	+10.29%
20	0.1954	0.1897	−2.92%	0.1783	0.1752	−1.75%	687.68	752.97	+9.49%
50	0.2573	0.2529	−1.70%	0.1923	0.1886	−1.95%	1690.62	1830.42	+8.27%

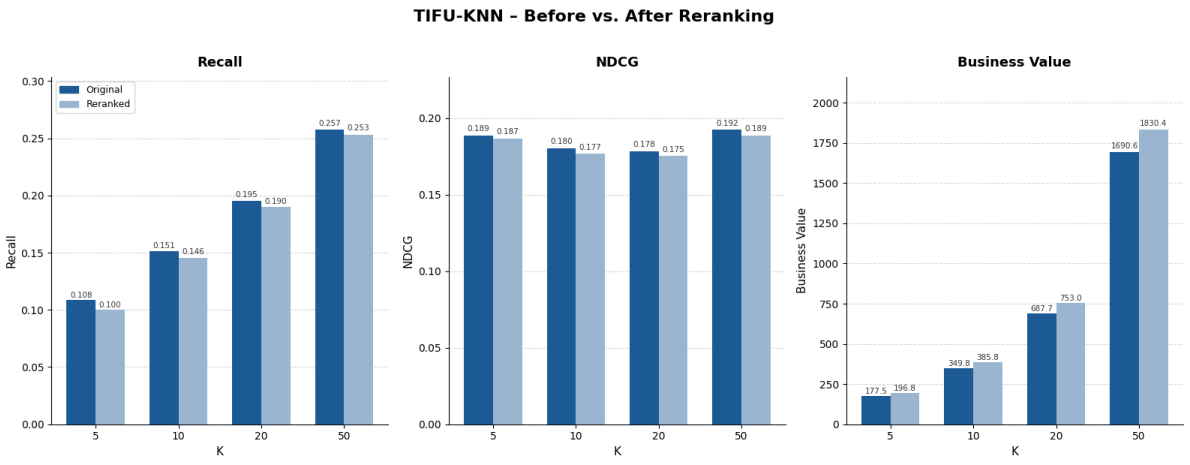

Figure 5.3: TIFU-KNN before and after multiplicative scoring reranking across all values of K. The left and centre panels show the modest accuracy losses in Recall and NDCG; the right panel shows the consistent BizValue gain.

BizValue improves by approximately 8 to 11% across all list sizes, with the largest gains observed at smaller K. This is expected as with a shorter recommendation list, each individual

item carries more weight in the aggregate score, and the reranking therefore has a greater relative impact. The accuracy losses are modest and diminish with increasing K. At $K = 50$, Recall drops by only 1.70%, which suggests that the two objectives are reasonably compatible at larger list sizes.

The NDCG losses remain consistently small (0.93% to 1.95%), indicating that the reranking does not substantially disturb the relative ordering of relevant items but rather introduces targeted swaps that place higher business value items at the expense of marginally less relevant ones.

5.2.2 ReCANet

The multiplicative scoring results for ReCANet follow a similar directional pattern. BizValue improves while Recall and NDCG decline. However, the trade-off is less favorable than for TIFU-KNN. The results are summarized in Table 5.6 and visualized in Figure 5.4.

Table 5.6: Performance of ReCANet under multiplicative scoring.

K	Rec (Orig)	Rec (Rer.)	Rec Δ%	NDCG (Orig)	NDCG (Rer.)	NDCG Δ%	Biz (Orig)	Biz (Rer.)	Biz Δ%
5	0.1215	0.1102	−9.32%	0.1980	0.1877	−5.23%	182.30	199.45	+9.41%
10	0.1711	0.1593	−6.93%	0.1952	0.1820	−6.77%	357.85	383.31	+7.11%
20	0.2150	0.2084	−3.07%	0.1955	0.1836	−6.12%	696.88	734.30	+5.37%
50	0.2773	0.2772	−0.04%	0.2101	0.2002	−4.70%	1641.52	1703.21	+3.76%

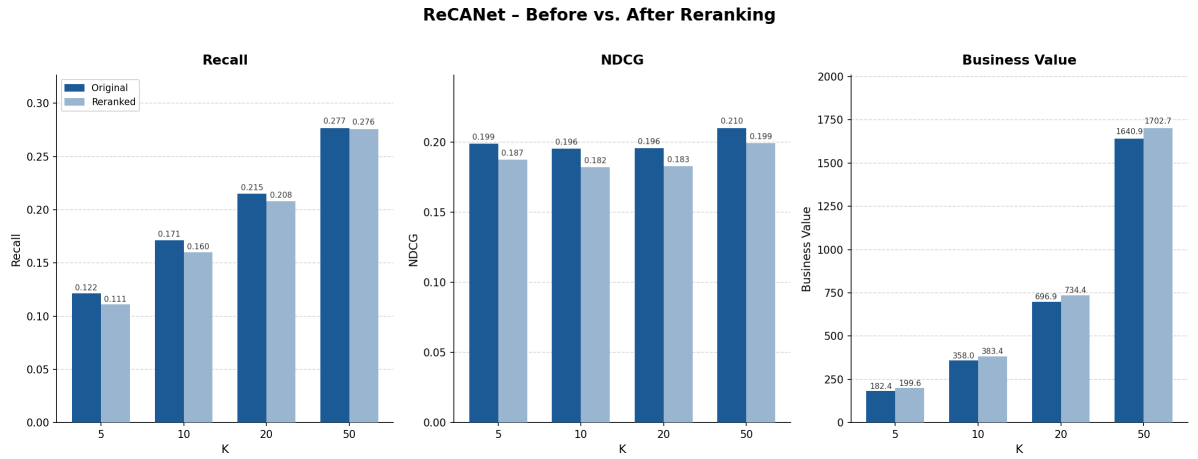

Figure 5.4: ReCANet before and after multiplicative scoring reranking across all values of K. Compared to TIFU-KNN, the NDCG losses in the centre panel are visibly larger while the BizValue gains in the right panel are smaller, illustrating the less favorable trade-off.

While BizValue improves across all K, the gains are noticeably smaller than those observed for TIFU-KNN ,approximately 4 to 9% compared to 8 to 11%. At the same time, the accuracy losses are comparable or larger, particularly for NDCG, which drops between 4.70% and 6.77% across most list sizes. A notable exception is $K = 50$, where Recall is virtually unchanged (−0.04%) while NDCG still drops by 4.70%. This can be attributed to the shallow score distribution at the lower end of ReCANet's candidate pool at that list length. Overall, ReCANet sacrifices more recommendation quality for a smaller business value gain, resulting in a less favorable trade-off.

5.2.3 Discussion

Multiplicative scoring successfully shifts both recommender systems toward higher business value recommendations. This demonstrates that even a simple post-processing approach can meaningfully incorporate business objectives without requiring architectural changes. However, the results reveal an important asymmetry between the two models. This asymmetry is directly visible in Figures 5.3 and 5.4. While the BizValue bars grow consistently for both models, the accuracy bars for ReCANet decrease noticeably more, particularly for NDCG, even though the BizValue gain is smaller.

TIFU-KNN responds more favorably to multiplicative reranking than ReCANet, achieving larger BizValue gains with smaller accuracy costs. This difference can be explained by examining the structural properties of the two models.

The most important factor lies in the composition of each model's candidate set. ReCANet is architecturally constrained to score only items that a user has previously purchased, treating recommendation as a binary classification problem over personal purchase history. While this design is effective for predicting repeat purchases, the resulting candidate pool is dominated by items that users buy routinely. Multiplicative scoring can only rerank within this already-constrained set, which limits the achievable BizValue improvement. TIFU-KNN, by contrast, incorporates a collaborative filtering component that draws on the purchase histories of similar users. This produces a more diverse candidate pool that includes items the target user has never purchased, and consequently contains a greater proportion of high-BizValue items that multiplicative scoring can elevate in the final ranking.

The candidate pool constraint also explains a notable anomaly at $K = 50$ for ReCANet. Recall drops by only 0.04%, yet NDCG still falls by 4.70%. At this list size, the pool of candidates with significant customer utility is nearly exhausted, leaving almost no room for multiplicative scoring to swap items into or out of the top-50 list. The negligible change in Recall reflects this. However, items that remain in the list are reordered, and since NDCG penalizes relevant items appearing at lower positions, the positional disruption is captured even when coverage remains unchanged.

A second contributing factor relates to the distribution of prediction scores produced by each model. ReCANet, as a neural network with sigmoid output, produces well-calibrated probabilities that tend to be sharply peaked. A small number of items receive high probabilities while the majority receive low ones. In such a distribution, the business score has limited leverage to reorder the top candidates, as the customer utility between them is large relative to the vari-

ation in business scores. The frequency-based scores of TIFU-KNN tend to be more uniformly distributed across candidates, giving the business score comparatively more influence over the final ranking.

These findings suggest that the effectiveness of multiplicative scoring is not solely a function of the reranking approach itself but is also shaped by the structural properties of the underlying recommender system. In particular, models that produce diverse candidate sets and relatively flat score distributions are more amenable to post-processing business-aware reranking than models that are optimized for precision on a narrow personal history.

Finally, multiplicative scoring provides no explicit control over the trade-off between recommendation accuracy and business value. The degree to which BizValue improves, and the corresponding accuracy cost, emerges from the interaction of score distributions rather than from a tunable parameter. This limitation motivates the optimization-based approach explored in the following section.

5.3 Mixed Integer Programming

The Mixed Integer Programming (MIP) approach represents a more principled method for business-aware reranking compared to multiplicative scoring. Rather than applying a fixed heuristic fusion, MIP explicitly models the recommendation task as a multi-objective optimization problem, allowing for provably optimal trade-offs between customer utility and business value. As described in Section 4.2, two formulations are evaluated: the weighted sum method and the epsilon-constraint method. Both are applied as post-processing steps to the same baseline models from Section 5.1, using the same dataset configuration of 1,500 users over a one-year window and the balanced business value score weighting throughout.

5.3.1 Weighted Sum

The weighted sum method combines customer utility and business value into a single objective through a scalar parameter $\lambda \in [0, 1]$, where higher values of λ place greater weight on predicted purchase probability and lower values shift emphasis toward business value. By systematically varying λ, the full trade-off frontier between the two objectives can be traced.

TIFU-KNN Figure 5.5 shows the parameter sensitivity heatmaps for TIFU-KNN under weighted sum reranking, displaying the percentage change in Recall, NDCG, and BizValue relative to the unmodified baseline across all evaluated values of λ and K.

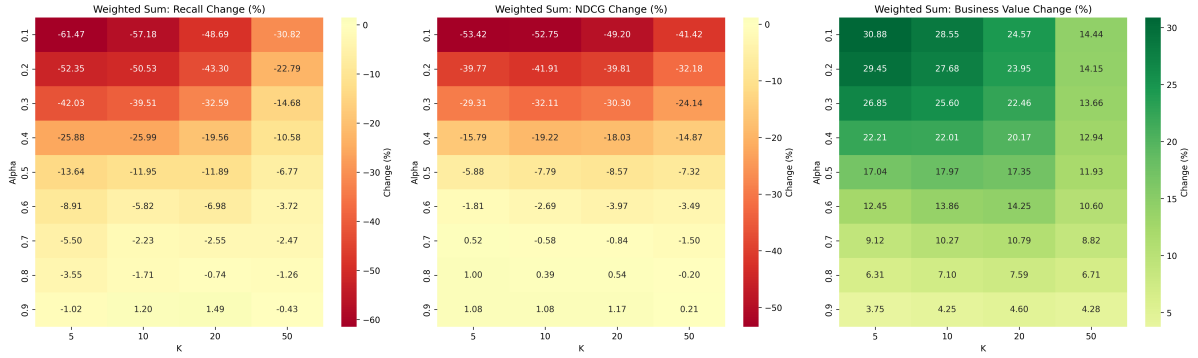

Figure 5.5: TIFU-KNN parameter sensitivity heatmaps for the weighted sum method. Each cell shows the percentage change relative to the unmodified baseline.

The results reveal a clear and well-behaved trade-off structure. At conservative settings ($\lambda = 0.8$), TIFU-KNN achieves BizValue improvements of approximately 6.3 to 7.6% across $K = 5$ to $K = 50$, while incurring only marginal accuracy losses. Recall drops by 0.7 to 3.6% and, notably, NDCG shows a slight positive change of $+1.0\%$ at $K = 5$ while remaining flat or slightly positive across $K = 10$ and $K = 20$. At $\lambda = 0.8$ the reranking improves business value without meaningful degradation in ranking quality and in some cases produces a marginal improvement. The NDCG gain is small and should be interpreted with caution rather than as a robust effect, but it suggests that at this parameter setting the weighted sum identifies recommendation lists that are simultaneously better for the business and at least as good for the customer in terms of ranked relevance. This constitutes a near-Pareto improvement and a particularly attractive operating point for practitioners.

At more aggressive settings, $\lambda = 0.7$ delivers BizValue gains of approximately 9 to 11% with Recall losses of 2 to 6% and NDCG losses of approximately 0 to 2%. This range represents the practical sweet spot for most operational contexts, providing meaningful business value improvement at an accuracy cost that remains modest and arguably acceptable given the commercial benefit.

At the extreme end, $\lambda = 0.2$ achieves BizValue gains of 14 to 29%, but at the cost of 23 to 52% Recall loss and 32 to 40% NDCG loss. Such a trade-off would be difficult to justify in production settings where customer satisfaction must be maintained.

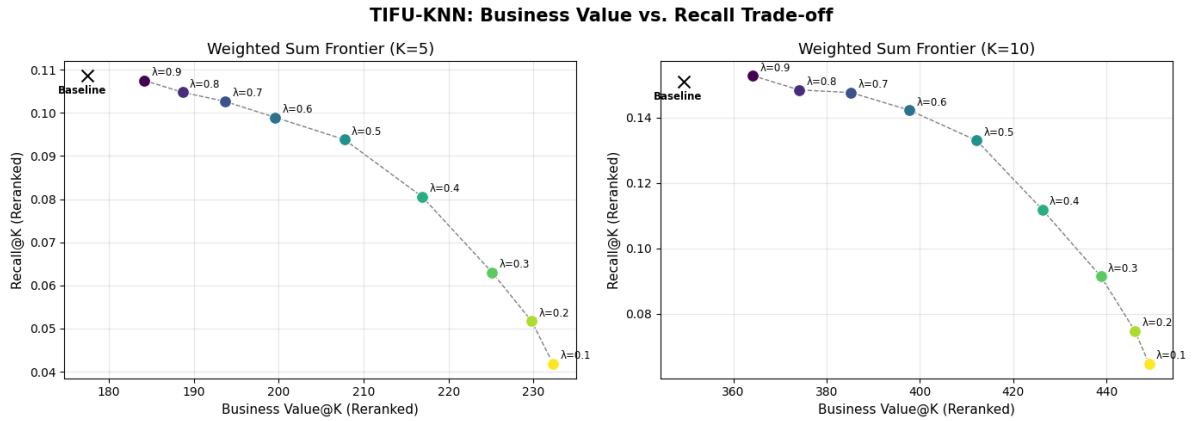

Figure 5.6: TIFU-KNN weighted sum Pareto frontier at $K = 5$ and $K = 10$. Each point corresponds to one value of λ; the baseline is marked with a cross. The frontier is well-spread and nearly linear, reflecting the flat score distribution of TIFU-KNN.

Figure 5.6 illustrates the Pareto frontier in BizValue-Recall space. The curve is well-spread across the full λ range. The points at $\lambda = 0.8$ and $\lambda = 0.9$ lie close to the baseline in recall while already delivering meaningful BizValue gains, confirming the near-Pareto character of the conservative operating region identified above.

ReCANet Figure 5.7 shows the corresponding heatmaps for ReCANet under weighted sum reranking.

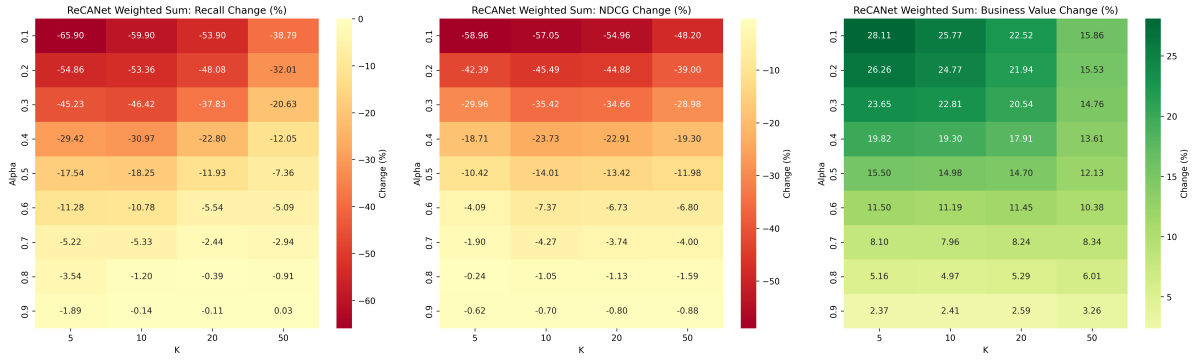

Figure 5.7: ReCANet parameter sensitivity heatmaps for the weighted sum method. Each cell shows the percentage change relative to the unmodified baseline.

ReCANet follows the same directional pattern but with consistently less favorable trade-offs. At $\lambda = 0.8$, BizValue improves by approximately 5.0 to 6.0%, while Recall drops by 0.4 to 3.5% and NDCG by 0.2 to 1.6%. At $\lambda = 0.7$, BizValue gains reach approximately 8.1 to 8.3% with Recall losses of 2.4 to 5.3% and NDCG losses of 1.9 to 4.3%. Compared to TIFU-KNN at the

same λ settings, ReCANet achieves similar BizValue gains at $K = 50$ but less favorable trade-offs at smaller list sizes, representing a consistently less efficient exchange across the parameter space.

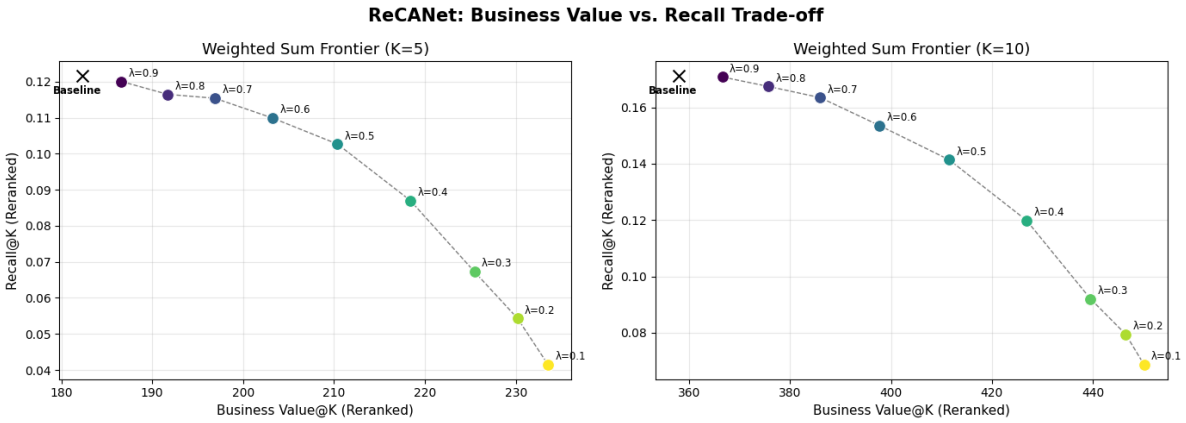

Figure 5.8: ReCANet weighted sum Pareto frontier at $K = 5$ and $K = 10$. Each point corresponds to one value of λ; the baseline is marked with a cross. Compared to TIFU-KNN, the frontier is similarly shaped but the achievable BizValue gains at moderate recall costs are smaller.

Figure 5.8 shows the corresponding frontier for ReCANet. The shape is qualitatively similar to that of TIFU-KNN, with a gentle slope at high λ values and a steeper drop-off as λ decreases. However, the frontier lies closer to the baseline in the BizValue dimension, which is consistent with the ceiling imposed by ReCANet's narrow candidate pool.

5.3.2 Epsilon-Constraint

The epsilon-constraint method takes a fundamentally different approach. Rather than blending objectives through a scalar weight, it maximizes customer utility subject to a hard lower bound on aggregate business value. The bound is defined as a percentile of the business value scores in the candidate set, so that $\varepsilon = 70$ requires the selected items to collectively meet at least the 70th percentile of available business value. This formulation is particularly suited to operational contexts where a minimum business value threshold must be guaranteed rather than merely encouraged.

TIFU-KNN An initial exploration across the full range of epsilon percentiles revealed that changes below the 70th percentile were negligible for TIFU-KNN. Accuracy metrics and BizValue remained essentially unchanged relative to the baseline, indicating that the constraint was not binding at lower thresholds. We therefore report results for the range $\varepsilon = 70$ to $\varepsilon = 90$, where meaningful trade-offs begin to emerge.

Figure 5.9 shows the epsilon-constraint heatmaps for TIFU-KNN.

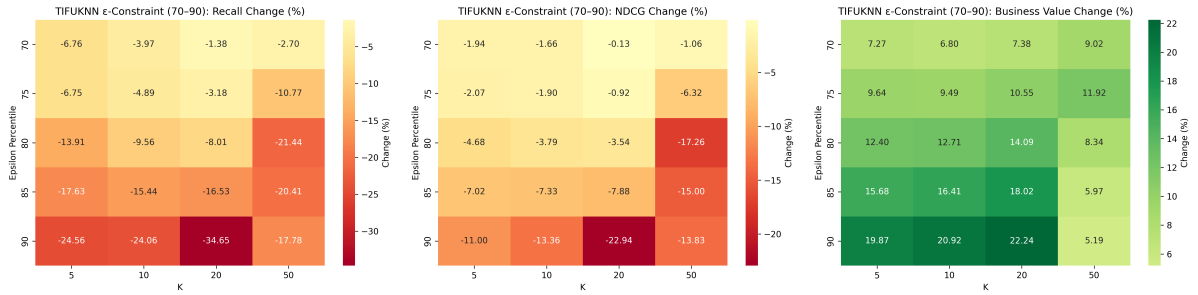

Figure 5.9: TIFU-KNN parameter sensitivity heatmaps for the epsilon-constraint method. Each cell shows the percentage change relative to the unmodified baseline.

At $\varepsilon = 70$, TIFU-KNN achieves BizValue gains of approximately 6.8 to 9.0% for $K = 5$, $K = 10$, and $K = 20$, with Recall losses of 1.4 to 6.8% and NDCG losses of 0.1 to 1.9%. At $\varepsilon = 80$, BizValue improvements reach 12.4 to 14.1% for the same values of K, while Recall drops by 8.0 to 13.9% and NDCG by 3.5 to 4.7%. As epsilon increases, both gains and costs escalate: at $\varepsilon = 90$, BizValue improvements reach 19.9 to 22.2% for $K = 5$, $K = 10$, and $K = 20$, while Recall drops by 24.6 to 34.7% and NDCG by 11.0 to 22.9%.

A notable anomaly appears at $K = 50$ across all epsilon levels, where BizValue gains collapse to approximately 5 to 9%, far below what is achieved at smaller K, while accuracy losses remain substantial. This behavior is consistent with a structural limitation of the epsilon-constraint formulation at large list sizes. When $K = 50$, the candidate pool is already exhausted down to items with very low predicted purchase probabilities. Forcing a high business value floor across all 50 slots requires the solver to select low-probability items that happen to carry high business scores, which destroys accuracy without generating proportionate BizValue gain since the absolute contribution of low-ranked items to aggregate BizValue is small. In some cases the constraint likely triggers the feasibility fallback described in Section 4.2, returning the probability-ranked baseline list and explaining the near-flat BizValue gain at $K = 50$ under high epsilon. This behavior confirms the theoretical prediction made in the methodology and represents an important practical limitation of epsilon-constraint reranking at large recommendation list sizes.

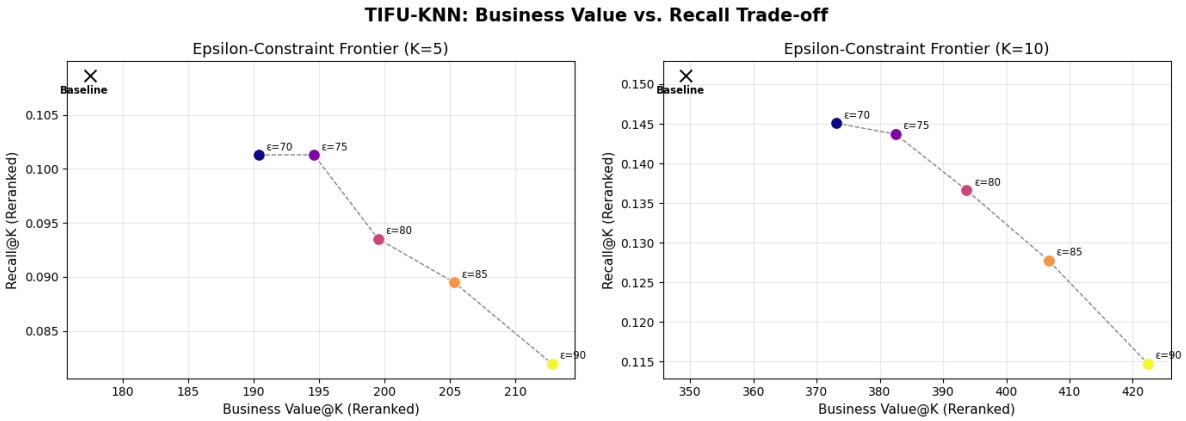

Figure 5.10: TIFU-KNN epsilon-constraint Pareto frontier at $K = 5$ and $K = 10$. Each point corresponds to one value of ε; the baseline is marked with a cross. The frontier is more concave than the weighted sum frontier, reflecting the harder constraint structure of the epsilon-constraint formulation.

Figure 5.10 shows the Pareto frontier for the epsilon-constraint method. Compared to the weighted sum frontier, the curve is noticeably more concave. Moderate epsilon values ($\varepsilon = 70$, $\varepsilon = 75$) achieve meaningful BizValue gains at relatively low recall cost, but the frontier bends sharply as epsilon increases toward 85 and 90, where each additional unit of BizValue gain requires a disproportionately large sacrifice in recall. This shape reflects the hard constraint structure of the formulation, where the solver must satisfy a strict floor rather than trading off objectives continuously.

ReCANet For ReCANet, the epsilon-constraint was evaluated over the same range as TIFU-KNN ($\varepsilon = 70$ to $\varepsilon = 90$). Like TIFU-KNN the constraint begins to produce meaningful trade-offs at $\varepsilon = 70$..

Figure 5.11 shows the epsilon-constraint heatmaps for ReCANet.

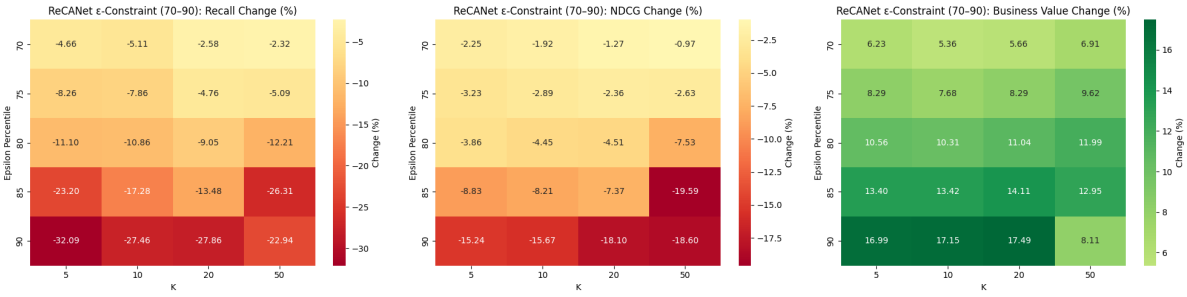

Figure 5.11: ReCANet parameter sensitivity heatmaps for the epsilon-constraint method. Each cell shows the percentage change relative to the unmodified baseline.

At $\varepsilon = 70$, BizValue gains reach approximately 5.4 to 6.9% at the cost of 2.3 to 5.1% Recall loss and 1.0 to 2.3% NDCG loss, indicating that the constraint is already actively shaping the recommendations at this threshold. At $\varepsilon = 75$, BizValue improves by approximately 7.7 to 9.6%, while Recall drops by 4.8 to 8.3% and NDCG by 2.4 to 3.2%. At $\varepsilon = 80$, BizValue improves by approximately 10.3 to 12.0% for $K = 5$ to $K = 50$, with Recall losses of 9.1 to 12.2% and NDCG losses of 3.9 to 7.5%. These trade-offs are less favorable than those achieved by TIFU-KNN at comparable BizValue gain levels.

The same $K = 50$ instability observed for TIFU-KNN appears here as well, and in fact more severely. At $\varepsilon = 85$, $K = 50$ produces a 26.3% Recall drop and a 19.6% NDCG loss for only 13.0% BizValue gain, a markedly worse ratio than at smaller list sizes for the same epsilon. At $\varepsilon = 90$, $K = 50$ yields only 8.1% BizValue gain despite a 22.9% Recall loss, compared to 17.0 to 17.5% BizValue gain at $K = 5$ to $K = 20$ for the same epsilon. This confirms that the $K = 50$ instability is not model-specific but a general structural limitation of epsilon-constraint reranking that stems from the exhaustion of high-probability candidates at large list sizes.

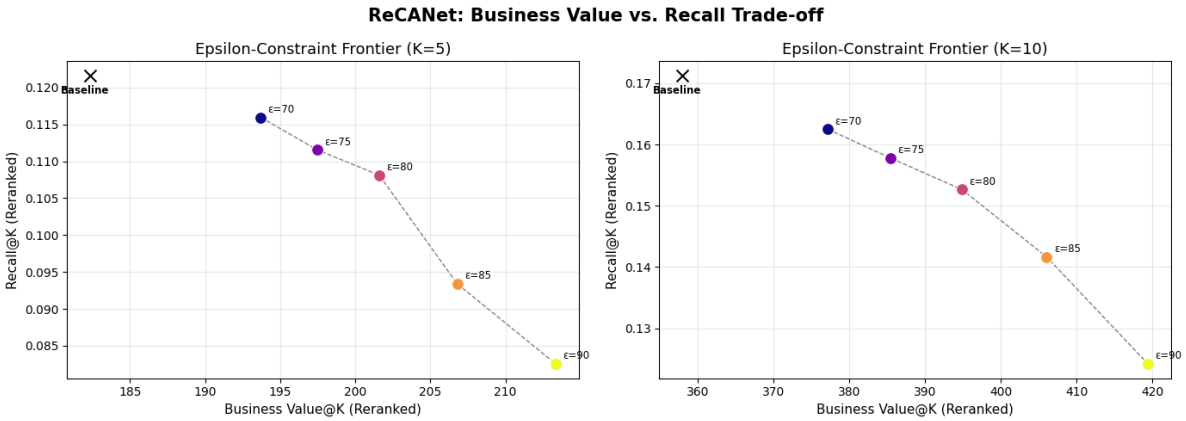

Figure 5.12: ReCANet epsilon-constraint Pareto frontier at $K = 5$ and $K = 10$. Each point corresponds to one value of ε; the baseline is marked with a cross. The frontier is steeper and more compressed than that of TIFU-KNN, reflecting the narrower range of achievable BizValue within ReCANet's candidate pool.

Figure 5.12 shows the corresponding frontier for ReCANet. Two differences from TIFU-KNN are visible. First, the BizValue axis spans a narrower range. Even at $\varepsilon = 90$ the achievable BizValue is lower than what TIFU-KNN reaches at $\varepsilon = 85$, again reflecting the ceiling imposed by the candidate pool composition. Second, the recall cost rises more steeply as epsilon increases, producing a sharper bend in the frontier. Together, these features confirm that epsilon-constraint reranking is structurally less effective for ReCANet than for TIFU-KNN, for the same reasons identified in the weighted sum analysis.

5.3.3 Discussion

Weighted Sum and Epsilon-Constraint as Complementary Tools The two MIP formulations serve as complementary tools suited to different operational contexts. The choice between them depends on the nature of the business objective being pursued.

The weighted sum is better suited to stable, day-to-day optimization where smooth and predictable control over the accuracy-business trade-off is desired. A retailer can set λ once and rely on consistent, proportional behavior. Increasing business emphasis by a fixed amount reliably produces a proportional BizValue gain and a predictable accuracy cost. This stability comes with an important caveat, however. The absolute trade-off achieved by a given λ depends on the distribution of business value scores in the candidate set, which can shift substantially from day to day as inventory levels change, products approach expiration, or promotions are activated. A λ that produced a balanced trade-off on one day may behave effectively as a customer-centric setting the next day if business scores have been compressed, or as an overly aggressive one if they have spread. In volatile operational environments, the weighted sum parameter therefore requires regular recalibration.

The epsilon-constraint sidesteps this problem by specifying the desired outcome directly rather than a mixing weight. A retailer that sets $\varepsilon = 80$ requires the recommendations to collectively meet at least the 80th percentile of available business value, regardless of how business scores are distributed on a given day. This formulation is robust to day-to-day score volatility because the constraint is defined relative to the current distribution rather than in absolute terms. Beyond daily volatility, the epsilon-constraint is the natural choice for hard operational requirements such as contractual commitments to promote a supplier's products, inventory liquidation targets for perishables approaching expiration, or regulatory obligations. These are fundamentally constraint-based requirements that a scalar weight cannot reliably satisfy. The cost of this robustness is a less predictable accuracy impact, particularly at extreme epsilon values and large K, where the instability identified above can lead to poor solutions.

In practice, a hybrid strategy may be most effective as using the weighted sum for routine daily operation and switching to the epsilon-constraint when specific hard business targets must be met.

Consistent Advantage of TIFU-KNN Under Post-Processing Reranking Across both MIP formulations and both evaluation metrics, TIFU-KNN responds more favorably to business-aware reranking than ReCANet. This finding is consistent with the pattern observed under multiplicative scoring in Section 5.2, and the two sections together present a coherent picture of why architecturally simpler models can be more amenable to post-processing optimization.

Three structural factors explain this difference. The most fundamental lies in candidate set composition. ReCANet is architecturally constrained to score only items from a user's personal purchase history, meaning its candidate pool is dominated by items the user repeatedly buys. The MIP optimizer can only select from within this constrained pool, which limits the ceiling of achievable BizValue improvement regardless of how aggressively the objective is weighted. TIFU-KNN's collaborative filtering component draws on the purchase histories of similar users, producing a broader and more diverse candidate pool that contains a greater proportion of high-BizValue items available for the optimizer to select.

The second factor concerns the variance of business value scores within each model's candidate pool. The effectiveness of MIP reranking scales with the degree of business value variation among candidates. If all items in the pool carry similar business scores, swapping one for another barely affects aggregate BizValue. The more diverse pool of TIFU-KNN exhibits higher business value variance, giving the optimizer meaningful degrees of freedom to construct substantially better-scoring recommendation lists.

Third, the frequency-based scores of TIFU-KNN are more uniformly distributed across candidates than the sigmoid-output probabilities of ReCANet, which tend to be sharply peaked around a small set of high-confidence items. For ReCANet, displacing the top-ranked items to satisfy business objectives incurs a steep accuracy penalty because the probability gap between top and lower-ranked candidates is large. The flatter score distribution of TIFU-KNN means that individual item swaps carry a smaller accuracy cost, allowing the optimizer to pursue more aggressive business-oriented reranking with less damage to recommendation quality.

What the MIP results add beyond the multiplicative scoring analysis is a demonstration that this advantage is not a coincidence of the specific heuristic applied in that earlier method, but a structural property that persists across optimization approaches. Even when the optimizer is given explicit and provably optimal control over the reranking, it cannot overcome the fundamental constraint that ReCANet's candidate pool is narrow and homogeneous in terms of business value. This observation has a broader implication for the design of business-aware recommender systems: post-processing reranking approaches are bounded by the quality and diversity of the candidate set they operate on. A model that produces a narrow, high-confidence candidate list optimized purely for personal repeat-purchase prediction will resist business-aware reranking regardless of the method applied. This motivates the model adjustment approaches explored in the following sections, where business objectives are integrated directly into the recommendation process rather than applied after the fact.

When Does MIP Add Value Over Multiplicative Scoring? The results across both MIP formulations raise a natural question for practitioners: given that multiplicative scoring requires no solver and can be implemented in two lines of code, when does the added complexity of MIP justify itself?

The answer depends on the magnitude of the business value target. At moderate BizValue gain levels of approximately 8 to 10%, the weighted sum method at $\lambda = 0.7$ to 0.8 achieves results broadly comparable to multiplicative scoring, offering little additional benefit beyond explicit parameter control. The MIP methods reach BizValue improvements above 15 up to 30% at $\varepsilon = 80$ to 90 for $K = 5$ to $K = 20$, a range clearly beyond what multiplicative scoring can deliver. MIP therefore adds clear value when strong business value targets must be met but offers diminishing returns at conservative settings where multiplicative scoring already performs well.

A second practical consideration concerns list size. The weighted sum remains well-behaved across all values of K, making it safe to deploy regardless of recommendation list length. The epsilon-constraint method, by contrast, degrades substantially at $K = 50$ across both models, as documented above. For deployments where long recommendation lists are required, the

weighted sum is the more reliable choice.

Finally, the Pareto frontier plots reveal a shape difference between the two formulations that carries operational meaning. The weighted sum frontier is approximately linear, meaning the trade-off between BizValue and recall is consistent across the full λ range and the parameter can be tuned incrementally. The epsilon-constraint frontier is concave, meaning moderate thresholds such as $\varepsilon = 70$ to 75 offer a disproportionately favorable trade-off relative to the accuracy cost, while aggressive thresholds yield rapidly diminishing BizValue returns per unit of recall sacrificed. When using the epsilon-constraint method, practitioners should therefore anchor their threshold in the moderate range unless a hard business requirement specifically demands a higher floor.

5.4 TIFU-KNN Model Adjustment Results

The business-aware decay extension modifies TIFU-KNN at the representation level rather than as a post-processing step. By blending temporal decay with item-level business weights through the parameter $\beta \in [0, 1]$, the modification reshapes every user vector in the model, affecting both the personal history component and the collaborative neighbor selection simultaneously.

5.4.1 Results

The results reveal a trade-off structure that is both well-behaved and notably favorable. At conservative settings ($\beta = 0.3$), the model achieves BizValue gains of approximately 3% at $K = 10$, while Recall and NDCG both remain within $\pm 0.3\%$ of the baseline. This constitutes a near-free improvement as the business-aware representation delivers meaningful commercial gain without degradation in recommendation quality.

As β increases, the trade-off becomes progressively more pronounced but remains gradual. At $\beta = 0.7$, BizValue improves by approximately 7%, while Recall and NDCG each decline by roughly 2 to 3%. At the extreme end, $\beta = 1.0$ achieves a BizValue gain of around 10% at a cost of approximately 6% Recall loss and 5% NDCG loss. Across all settings, $K = 50$ consistently shows the smallest accuracy penalties. At $\beta = 0.7$, Recall at $K = 50$ is essentially flat while still accumulating meaningful BizValue gains, reflecting the dilution effect at larger list sizes discussed in Section 5.2.

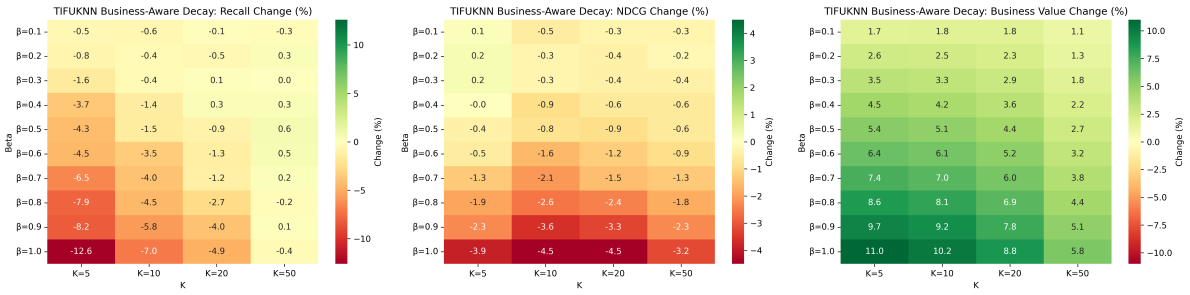

Figure 5.13: Percentage change in Recall, NDCG, and Business Value relative to the baseline, across all β values and list sizes $K \in \{5, 10, 20, 50\}$.

Figure 5.13 provides a comprehensive view of the trade-off landscape across the full $\beta \times K$ space. The three panels exhibit a characteristic asymmetry that reflects the nature of the trade-off. The Business Value heatmap is uniformly positive, with gains ranging from approximately 1% at low β and large K to over 11% at $\beta = 1.0$ and $K = 5$. The Recall and NDCG panels, by contrast, are predominantly near-neutral at low β, with notable losses appearing only from $\beta = 0.7$ onward and primarily at smaller list sizes.

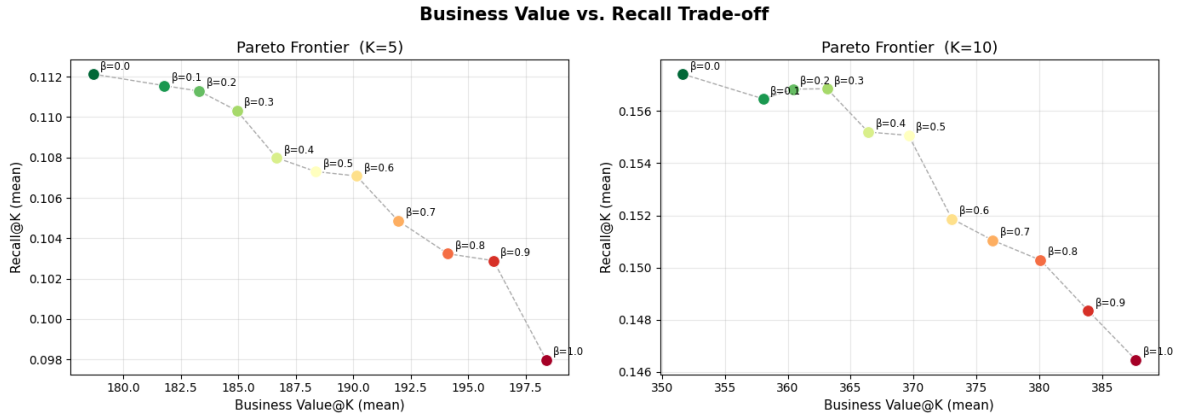

Figure 5.14: Pareto frontier of the business value and recall trade-off at $K = 5$ and $K = 10$.

The Pareto frontier in Figure 5.14 confirms this pattern. The curve slopes gently in the low-β region, meaning small movements along it yield substantial BizValue gains for only modest recall sacrifices.

5.4.2 Discussion

The TIFU-KNN model adjustment achieves a more favorable trade-off profile than either multiplicative scoring or MIP-based reranking, particularly at conservative and moderate β values. However, the effects are on a lower scale. At low settings, the approach produces near-Pareto

improvements that none of the post-processing methods evaluated in this work could match at comparable BizValue gain levels. Post-processing approaches consistently required trading accuracy for business value from the moment any business signal was introduced.

This difference can be attributed to the mechanism through which business awareness operates. Post-processing reranking acts on a fixed candidate set generated without any consideration of business objectives, so any business-oriented reordering necessarily displaces items that the base model ranked highly for relevance. The model adjustment, by contrast, reshapes the user representation before any scoring takes place. High-value items receive slightly amplified weight in every user's history vector, making them more likely to emerge naturally near the top of an already relevance-aware ranking. The business signal does not override the relevance signal but is woven into the same representation from which relevance is derived. This structural integration explains why accuracy costs are gradual and smooth rather than exhibiting the steeper, more immediate penalties that characterize post-processing approaches.

A second contributing factor is the propagation of the business signal through both the personal and collaborative components simultaneously. Because the KDTree used for neighbor selection is built from the same business-inflected vectors, the collaborative part of the prediction is also drawn toward users who share preferences for high-value items. This dual-channel amplification increases the effective influence of β without requiring extreme parameter values to achieve meaningful commercial impact.

Overall, the TIFU-KNN model adjustment provides a clear empirical answer for this particular model and integration method. Business value can be improved meaningfully and with fine-grained control, at accuracy costs that are negligible at conservative settings and modest at moderate ones. The smooth, monotonic response of all three metrics across the full β range makes the system straightforward to calibrate for operational deployment.

5.5 ReCANet Adjustment

The ReCANet adjustment integrates business awareness directly into the model's training process through exponential loss weighting. As described in Section 5.5, each training sample is assigned a weight

$$w_i = \exp\Big(\lambda \cdot \tilde{b}_i\Big), \tag{5.1}$$

where $\tilde{b}_i \in [0, 1]$ is the min-max normalized business value score of item i and λ is the weighting parameter. Positive values of λ upweight high-value items during training, causing the optimizer to allocate larger gradient updates to commercially important predictions. At $\lambda = 0$, all samples receive equal weight and the model reduces to the unmodified baseline. The parameter is varied across a grid spanning $\lambda \in \{+1, +2, +3, +4, +5\}$ to characterize the response of the weighting scheme across increasing levels of business-aware upweighting.

5.5.1 Results

Figure 5.15 presents the parameter sensitivity heatmap for ReCANet under exponential loss weighting, showing the percentage change in Recall, NDCG, and BizValue relative to the un-

modified baseline across all evaluated values of λ and K.

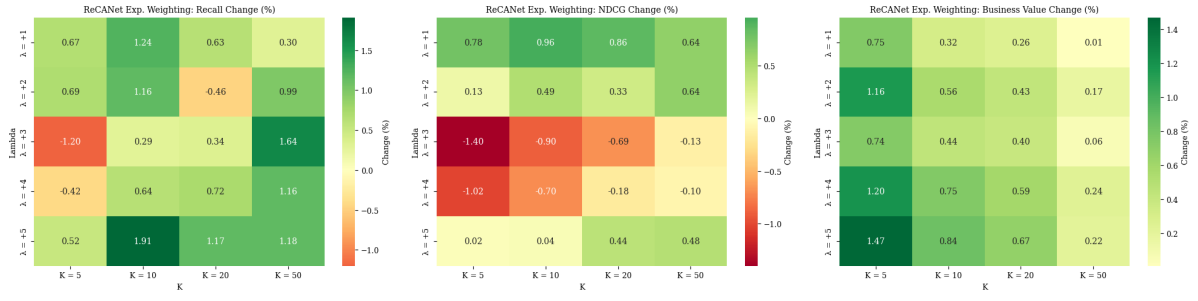

Figure 5.15: ReCANet parameter sensitivity heatmaps for the exponential loss weighting method. Each cell shows the percentage change relative to the unmodified baseline ($\lambda = 0$).

The most notable characteristic of the BizValue heatmap is not the direction of change but its magnitude. While BizValue improves across all 20 evaluated λ-K combinations without exception, the gains are small, ranging from $+0.01\%$ at $\lambda = +1$, $K = 50$ to $+1.47\%$ at $\lambda = +5$, $K = 5$. Compared to the BizValue improvements achieved by post-processing approaches applied to ReCANet, which reached approximately 4 to 9% under multiplicative scoring, the training-time intervention produces a considerably more modest commercial uplift. The weighting scheme operates in the intended direction, but within tight bounds.

The more distinctive finding concerns Recall and NDCG. Across the evaluated parameter range, both accuracy metrics remain largely stable or positive. At $\lambda \in \{+1, +2\}$, Recall ranges from $+0.67\%$ to $+1.24\%$ and from $+0.69\%$ to $+0.99\%$ respectively across K values, while NDCG follows a comparable pattern. This stands in contrast to the accuracy behavior observed under post-processing for ReCANet, where multiplicative scoring and MIP-based reranking produced Recall losses of up to 9.32% and NDCG declines of up to 6.77%. Exponential loss weighting achieves its BizValue gains without imposing the accuracy penalties that accompanied those approaches.

At $\lambda \in \{+3, +4\}$, accuracy metrics deteriorate. The largest Recall loss is -1.20% at $\lambda = +3$, $K = 5$, and the largest NDCG decline is -1.40% at the same setting. These losses are not accompanied by commensurate BizValue gains: $\lambda = +3$, $K = 5$ yields only $+0.74\%$ in BizValue, a figure comparable to what is achieved at lower λ values without any accuracy cost. This intermediate zone is therefore the least favorable region of the parameter space.

At $\lambda = +5$, the pattern reverses. Recall recovers, with gains of $+0.52\%$, $+1.91\%$, $+1.17\%$, and $+1.18\%$ across $K \in \{5, 10, 20, 50\}$, and NDCG stabilizes near zero or turns positive. BizValue simultaneously reaches its highest observed values across the heatmap. The $\lambda = +5$ configuration is the only setting that combines non-negative accuracy performance across all list sizes with the largest observed commercial uplift, making it the dominant configuration within the evaluated grid.

5.5.2 Discussion

The small magnitude of BizValue gains under exponential loss weighting points directly to the structural constraint identified in earlier sections. ReCANet's candidate pool at inference time consists exclusively of items each user has previously purchased. Loss weighting reshapes the gradient signal during backpropagation and adjusts how the model scores items within this pool, but it cannot introduce items that were never purchased by a given user. The ceiling on achievable BizValue improvement is therefore determined by the composition of the candidate pool, and no training-time intervention operating within this architecture can raise it substantially. This explains why the gains, while consistent in direction, remain small in absolute terms regardless of how aggressively λ is increased.

The more noteworthy result is the behavior of Recall and NDCG. Post-processing approaches applied to ReCANet consistently sacrificed accuracy for commercial gain. Multiplicative scoring reduced Recall by up to 9.32% and NDCG by up to 6.77%, while MIP-based reranking produced comparable penalties. Exponential loss weighting breaks this pattern. By operating at training time rather than as a post-hoc reranking step, the adjustment integrates the business signal into the model's learned representations rather than overriding its relevance-based ordering at inference. As a result, the model's ranking calibration is preserved in a way that post-processing approaches cannot achieve. This makes exponential loss weighting a qualitatively different integration mode for ReCANet as it yields smaller BizValue gains than post-processing but does so without the accuracy cost that has accompanied every other integration method evaluated for this model.

The non-monotonic response across λ values adds a practical dimension to these findings. The accuracy deterioration at $\lambda \in \{+3, +4\}$, followed by recovery at $\lambda = +5$, suggests that intermediate weighting strengths can destabilize the model's learned ranking without yet producing a coherent business-aware reorientation. Once the signal is strong enough at $\lambda = +5$ to dominate the loss landscape, the model adapts more consistently. Practitioners should therefore not assume a linear relationship between λ and outcome quality.

Taken together, these results refine the answer to Research Question 3: *how do the structural properties of a recommender system influence its responsiveness to business-aware optimization?* For ReCANet, the candidate pool constraint caps the commercial uplift achievable through any integration method, regardless of whether it operates at the training, scoring, or reranking stage. Within this constrained ceiling, however, training-time integration via exponential loss weighting emerges as the preferable approach. It is the only method evaluated for ReCANet that does not require trading recommendation quality for business value. The implication for practitioners is that when working with repeat-purchase architectures, training-time business-aware integration should be prioritized over post-processing, not because it produces larger commercial gains, but because it preserves the accuracy that makes the underlying model useful.

5.6 Customer-Level Impact Analysis

The preceding sections evaluated each business-aware integration method in terms of aggregate performance metrics averaged across all users. While these averages establish the overall trade-off profiles of each approach, they do not reveal whether the gains and losses are distributed evenly across the customer base or concentrated among specific customer types. A retailer deploying any of these methods in production needs to know not only the expected average effect but also which customers are likely to benefit, which are likely to experience degraded recommendations, and what observable behavioral features predict these outcomes.

This section addresses these questions through a customer-level segmentation analysis. For each of the four business-aware approaches (multiplicative scoring, weighted sum reranking, epsilon-constraint reranking, and the respective model adjustment), every user is scored under both the baseline and the modified method. The per-user change in relevance (Δ_{rel}, defined as the average of the change in Recall@K and NDCG@K) and the per-user change in aggregate business value (Δ_{biz}) are then computed.

Parameter selection. Each approach is evaluated at the parameter setting identified in Sections 5.2 to 5.5 as offering the most favorable trade-off between business value gain and accuracy cost. The weighted sum uses $\lambda = 0.6$, which falls within the practical sweet spot ($\lambda = 0.6$ to 0.7) identified in Section 5.3.1, delivering BizValue gains of approximately 9 to 11% with Recall losses of 2 to 6% and NDCG losses of approximately 0 to 2% for TIFU-KNN. The epsilon-constraint uses the 70th percentile threshold, which Section 5.3.2 identified as the onset of meaningful trade-offs for TIFU-KNN and an already binding constraint for ReCANet, offering a BizValue improvement of approximately 7 to 9% at modest accuracy cost. For the model adjustments, TIFU-KNN uses $\beta = 0.3$, the conservative setting that Section 5.4 showed achieves approximately 3% BizValue gain with Recall and NDCG both within $\pm 0.3\%$ of the baseline. ReCANet uses exponential loss weighting with $\lambda_{\text{exp}} = 5$, the strongest positive weighting evaluated in Section 5.5. Although that section demonstrated that this adjustment produces negligible aggregate effects due to ReCANet's fixed candidate pool, it is included here to examine whether the customer-level distribution reveals heterogeneous effects that the aggregate averages might mask. Multiplicative scoring, having no tunable parameter, is applied as described in Section 5.2.

Choice of $K = 10$. The analysis is conducted at $K = 10$. This list size is chosen for three reasons. First, it represents the most operationally relevant setting for grocery recommendation, where displayed recommendation lists are typically short. Second, the parameter sensitivity analyses in Sections 5.2 to 5.5 show that the trade-off behavior at $K = 10$ is representative of the broader pattern across list sizes. It avoids both the high variance of $K = 5$, where individual item swaps have outsized effects on per-user metrics, and the dilution effects at $K = 50$, where the epsilon-constraint method degrades due to candidate pool exhaustion. Third, $K = 10$ is the standard evaluation point used throughout this work for detailed analysis, making results directly comparable with the aggregate findings reported in earlier sections.

Users are assigned to one of three segments based on threshold rules applied to the per-user deltas:

- **Good**: both customer relevance and business value improve ($\Delta_{\text{rel}} > 0$ and $\Delta_{\text{biz}} > 0$).
- **Bad**: customer relevance degrades meaningfully while business value gains are modest or negative ($\Delta_{\text{rel}} < -0.25$, or $\Delta_{\text{biz}} < 0$, or both $\Delta_{\text{biz}} \leq 15$ and $\Delta_{\text{rel}} < 0$).
- **Balanced**: all remaining users, where the trade-off is moderate or neutral.

Users for whom both the baseline and modified method produce zero relevance (zero Recall and zero NDCG under both conditions) are excluded from the analysis, as no meaningful comparison can be drawn for these cases. To characterize each segment, a set of behavioral features is computed from each user's transaction history. Number of baskets, average basket size, total purchases, number of unique items, item diversity (unique items divided by total purchases), repeat purchase rate, average historical business value score of purchased items, and ground-truth set size. A Random Forest classifier trained on these features identifies which behavioral dimensions are most predictive of segment membership.

5.6.1 TIFU-KNN

Figure 5.16 presents the cross-approach comparison for TIFU-KNN across all four integration methods. The scatter plots in the top row show the joint distribution of Δ_{rel} and Δ_{biz} for each user, colored by segment assignment. The middle row reports segment sizes, and the bottom row shows Random Forest feature importances for predicting segment membership.

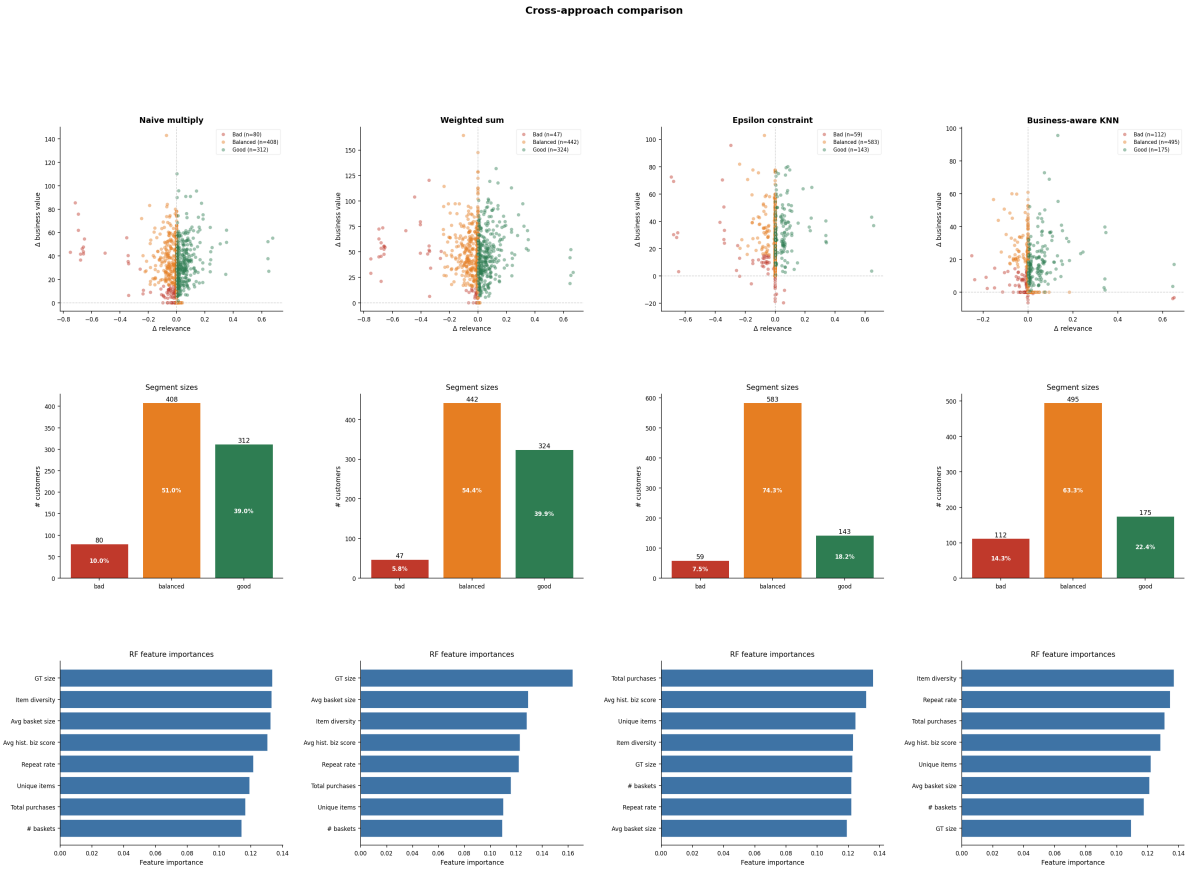

Figure 5.16: TIFU-KNN cross-approach comparison. Top row: per-user scatter of relevance change and business value change, colored by segment. Middle row: segment size distributions. Bottom row: Random Forest feature importances for predicting segment membership.

Segment Distributions The four approaches produce noticeably different segment distributions. The weighted sum achieves the lowest proportion of negatively affected customers (5.8%, $n = 47$), followed by the epsilon-constraint method (7.5%, $n = 59$), multiplicative scoring (10.0%, $n = 80$), and the business-aware TIFU-KNN adjustment (14.3%, $n = 112$). The epsilon-constraint pushes the largest share of users into the balanced segment (74.3%, $n = 583$), which is consistent with its conservative, floor-based formulation. Most users experience a business value improvement without a relevance loss large enough to trigger the bad threshold, but also without a simultaneous relevance gain that would qualify them as good. The business-aware TIFU-KNN, by contrast, produces the most polarized distribution, with a larger bad segment but also a meaningful good segment (22.4%, $n = 175$).

The scatter plots reveal a structural difference between post-processing and model-level integration. For the three post-processing approaches (multiplicative scoring, weighted sum, epsilon-constraint), the scatter clouds are concentrated in the upper-left and upper-right quadrants, with Δ_{biz} predominantly positive and Δ_{rel} spanning a wide range. This is expected, as

reranking can only redistribute items within the candidate set, so it reliably increases business value but introduces variable relevance effects depending on the extent of reordering required. The business-aware TIFU-KNN scatter, by contrast, is more tightly centered around the origin, with both positive and negative Δ_{biz} values. This reflects the more subtle mechanism of the model adjustment as it reshapes user representations rather than reordering candidates, producing smaller but more symmetric perturbations.

Behavioral Profiles Table 5.7 reports the median behavioral feature values for each segment across all four approaches.

Table 5.7: TIFU-KNN segment median values across approaches.

	Mult. scoring			Weighted sum			Epsilon constr.			Biz-aware TIFU-KNN		
Feature	Bad	Bal.	Good	Bad	Bal.	Good	Bad	Bal.	Good	Bad	Bal.	Good
# baskets	36	42	40	38	42	40	26	42	42	52	42	33
Avg basket size	9.9	9.7	10.3	7.8	9.7	10.4	10.3	9.9	10.6	11.0	9.7	10.5
Total purchases	383	441	411	428	416	411	332	437	435	514	428	332
Unique items	253	300	275	255	288	280	221	295	258	330	286	237
Diversity	0.66	0.69	0.70	0.63	0.68	0.70	0.74	0.68	0.70	0.62	0.69	0.73
Repeat rate	0.23	0.21	0.21	0.25	0.21	0.21	0.19	0.21	0.22	0.25	0.21	0.19
Avg hist. biz	33.9	33.8	34.0	33.9	33.8	34.0	34.0	33.9	33.5	33.7	33.9	33.9
GT size	6	10	13	2	10	13	9	10	14	12	9	14

Across every approach, the Random Forest classifier identifies ground-truth basket size as the most important predictor of segment membership. The effect is most pronounced for the weighted sum, where the bad segment has a median GT size of only 2 items compared to 13 for the good segment.

GT size is not a customer characteristic but a property of a single held-out shopping occasion, and its dominance reflects two structural properties of the evaluation. Larger ground-truth baskets give any recommender more opportunities to register a hit, raising baseline accuracy before any reranking is applied. They also dilute the cost of individual item displacements as removing one relevant item from a two-item basket halves Recall, while the same removal from a fifteen-item basket reduces it by only a few percentage points. A regular customer with a stable purchase pattern can therefore land in the bad segment simply because the held-out trip happened to be a small top-up visit. The GT size signal speaks to the structure of the evaluation instance rather than to any behavioral trait that distinguishes amenable from vulnerable customers.

The customer-level interpretation therefore shifts to features ranked second and third, which describe the customer rather than the test instance. For the three post-processing approaches, item diversity emerges as the leading behavioral predictor. The bad segment consistently exhibits lower diversity than the good segment. Customers with low diversity have purchase histories concentrated on a small set of repeatedly bought products, producing sharp, confident base-model predictions that the reranker cannot disturb without cost. Customers with high

diversity have flatter score distributions, giving the optimizer multiple near-equivalent items to select from when satisfying the business objective.

Average historical business value score adds a second dimension. Although the absolute differences across segments are small (medians 33.7 to 34.0), its consistent appearance in the top features indicates that customers whose existing patterns already align with high-business-value items absorb reranking more comfortably, since fewer genuinely novel substitutions are required.

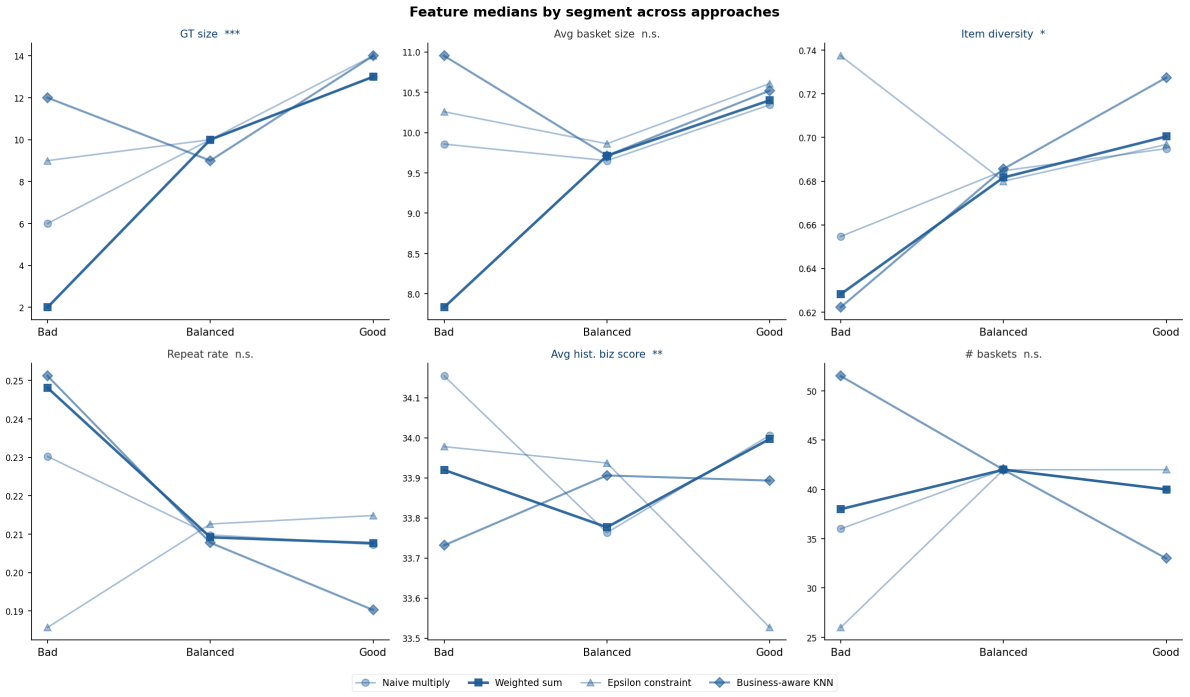

Figure 5.17: TIFU-KNN feature medians by segment across all four approaches, with Kruskal-Wallis significance annotations. Each line represents one approach; colored dots indicate segment identity.

The business-aware TIFU-KNN adjustment exhibits an inverted pattern. Its bad segment contains the most active customers in the dataset: highest baskets (52), highest total purchases (514), lowest diversity (0.62), and highest repeat rate (0.25). These habitual shoppers have routinized baskets that the business-aware representation disrupts by amplifying items they do not prefer. The good segment contains lighter, more exploratory shoppers (33 baskets, diversity 0.73, repeat rate 0.19) whose purchasing is not dominated by a fixed staple set and therefore absorbs the reweighting without major disruption.

This inversion is interpretively important. Where post-processing results are partly driven by the test-instance effect captured by GT size, the model-adjustment results respond to genuine and stable customer traits. Post-processing displaces items from a fixed ranking, harming users with few relevant items to spare on a given trip. Model adjustment reshapes the user representation itself, harming users whose established routines are most rigidly encoded in it. The two

paradigms therefore expose different populations to risk.

5.6.2 ReCANet

Figure 5.18 presents the corresponding analysis for ReCANet. The four approaches are multiplicative scoring, weighted sum ($\lambda = 0.6$), epsilon-constraint ($\varepsilon = 70$), and the exponential loss-weighted model ($\lambda_{\text{exp}} = 5$).

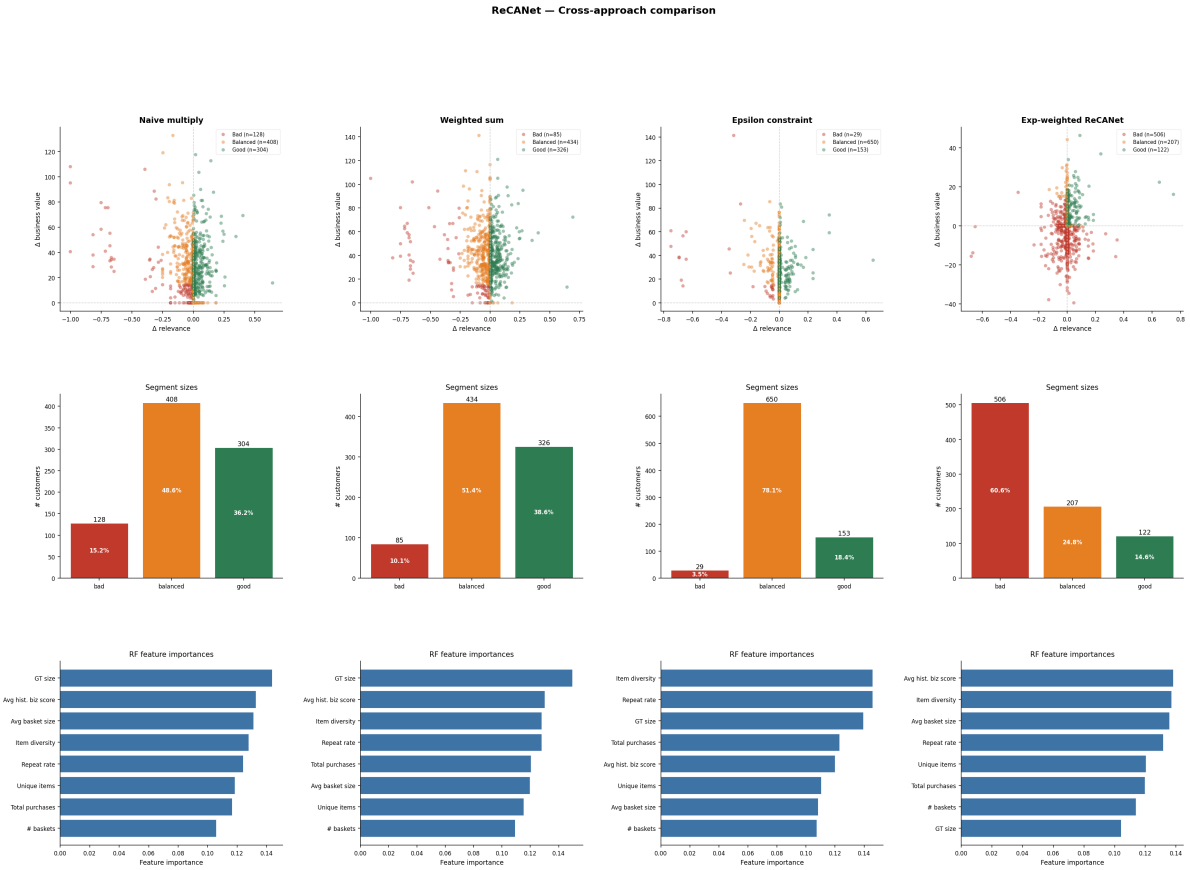

Figure 5.18: ReCANet cross-approach comparison. Layout identical to Figure 5.16.

Segment Distributions The segment distributions for ReCANet's three post-processing approaches follow the same qualitative pattern as TIFU-KNN but with consistently larger bad segments. Multiplicative scoring produces 15.2% bad ($n = 128$), weighted sum 10.1% ($n = 85$), and epsilon-constraint 3.5% ($n = 29$). The epsilon-constraint result is notable as it achieves the smallest bad segment of any approach on either model, pushing 78.1% of users into balanced and 18.4% into good. This reflects the conservative nature of a floor-based constraint applied to ReCANet's already narrow candidate pool, where limited room for aggressive reranking means that most users experience only modest perturbations.

The exp-weighted ReCANet presents a qualitatively different picture. Its scatter plot is the most striking departure from the other three: Δ_{biz} is centered near zero and spans both positive and negative values (approximately -40 to $+50$), while Δ_{rel} shows substantial spread. This produces the largest bad segment of any approach on either model: 60.6% ($n = 506$), with only 14.6% good ($n = 122$) and 24.8% balanced ($n = 207$). The interpretation is consistent with the aggregate findings of Section 5.5: exponential loss weighting does not meaningfully shift the business value distribution of ReCANet's recommendations because the candidate pool is fixed at inference time. This illustrates why aggregate metrics can mask distributional harm. The small positive shift reported in Section 5.5 coexists with a majority of users experiencing relevance degradation uncompensated by business value gains. The aggregate uplift is driven by a minority of users in the positive tail of Δ_{biz}.

Behavioral Profiles Table 5.8 reports the median behavioral features by segment for ReCANet.

Table 5.8: ReCANet segment median values across approaches.

	Mult. scoring			Weighted sum			Epsilon constr.			Exp-wt. ReCANet		
Feature	Bad	Bal.	Good	Bad	Bal.	Good	Bad	Bal.	Good	Bad	Bal.	Good
# baskets	47	42	41	54	41	40	64	40	48	43	42	37
Avg basket size	10.0	9.4	11.0	10.5	9.7	10.7	10.9	9.7	11.2	10.7	9.1	9.6
Total purchases	442	390	477	530	397	440	750	389	541	461	359	365
Unique items	284	272	340	333	273	307	388	271	370	309	252	254
Diversity	0.65	0.69	0.69	0.62	0.70	0.69	0.56	0.70	0.66	0.69	0.68	0.71
Repeat rate	0.21	0.21	0.21	0.26	0.20	0.21	0.27	0.20	0.23	0.21	0.21	0.20
Avg hist. biz	33.9	33.9	33.9	33.9	33.8	33.9	33.9	33.9	33.7	33.9	34.0	33.9
GT size	6	10	12	3	10	12	9	10	14	11	8	11

The GT size pattern replicates for ReCANet, with the bad segment showing the smallest median GT size (3 to 9 across post-processing approaches) and the good segment the largest (12 to 14). As with TIFU-KNN, this reflects the structure of the evaluation instance rather than a behavioral trait, and the customer-level interpretation rests on the second- and third-ranked features.

For ReCANet, the leading behavioral feature is average basket size. This contrasts with TIFU-KNN, where item diversity played that role, and the difference is consistent with the distinct candidate generation mechanisms of the two models. ReCANet scores only previously purchased items, so the relevant behavioral dimension is how many items the user typically buys per trip. Customers who routinely fill large baskets generate denser candidate pools where reranking has more room to operate. Customers with consistently small baskets have shallow pools where each item is harder to replace without quality loss. Crucially, average basket size is a stable customer trait observable from purchase history, and it captures much of the same predictive signal as the test-instance GT size in a behaviorally interpretable form.

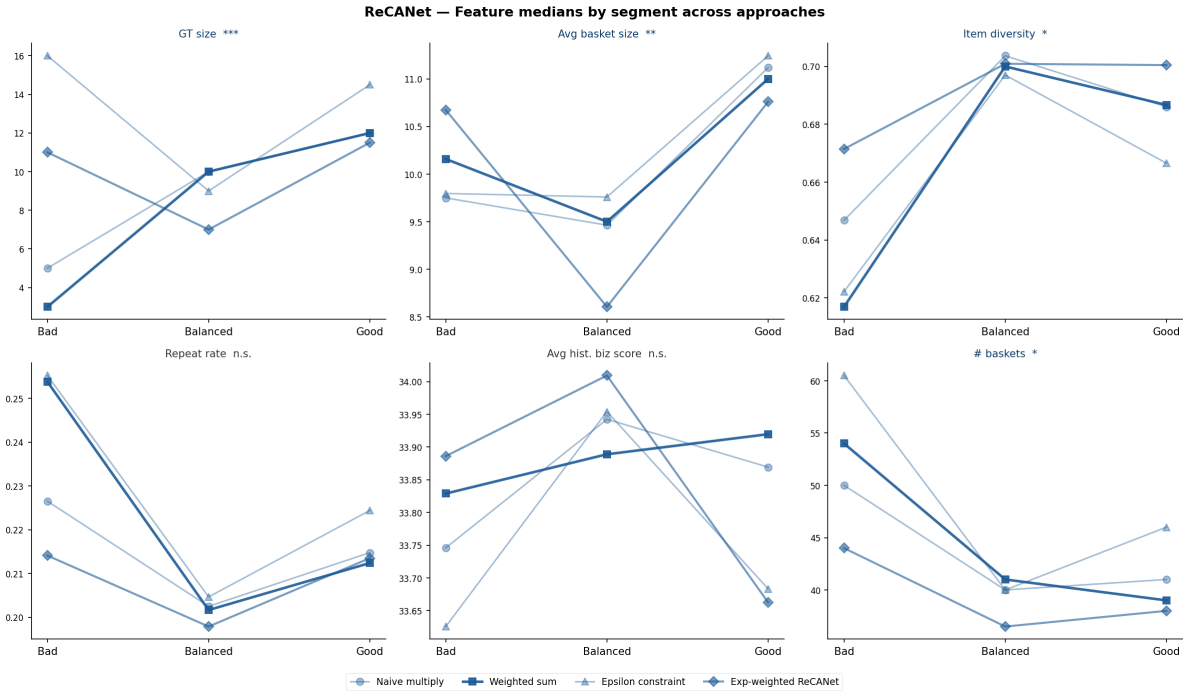

Figure 5.19: ReCANet feature medians by segment across all four approaches, with Kruskal-Wallis significance annotations.

The epsilon-constraint approach produces a particularly informative bad segment: 29 users with the highest median number of baskets, most total purchases, and lowest diversity of any segment in the analysis. These extreme habitual shoppers cannot absorb the item substitutions forced by the hard business value floor without substantial relevance loss, mirroring the inversion observed for the business-aware TIFU-KNN adjustment.

The exp-weighted ReCANet stands apart. GT size does not discriminate between segments (medians of 11, 8, and 11), and no single feature dominates the importance ranking. This is consistent with the aggregate finding from Section 5.5. Because the candidate pool is fixed at inference time, the perturbations introduced by loss weighting are not systematically related to observable customer features and behave more like training-time noise than a coherent behavioral reorientation.

5.6.3 Cross-Model Comparison

Comparing the customer-level results across the two base models reveals both robust patterns and informative divergences.

Consistent findings. Three results hold across both models and all post-processing approaches. First, ground-truth basket size is the single strongest statistical predictor of segment membership across both models and all post-processing approaches, but this signal reflects properties

of the evaluation instance rather than of the customer. The behaviorally meaningful predictors that emerge once GT size is set aside, item diversity for TIFU-KNN and average basket size for ReCANet, point to a consistent underlying mechanism. Customers whose purchasing leaves the recommender with a broader candidate set are better fitted for business-aware reranking than customers whose purchasing produces narrow, high-confidence predictions that the reranker cannot disturb without cost. Second, the weighted sum consistently produces the smallest or near-smallest bad segment among the post-processing approaches (5.8% for TIFU-KNN, 10.1% for ReCANet), confirming its favorable risk profile at the customer level as well as in aggregate. Third, the epsilon-constraint method produces the highest proportion of balanced-segment users (74.3% for TIFU-KNN, 78.1% for ReCANet), reflecting its conservative floor-based design that avoids both large gains and large losses for most customers.

Model-specific differences. ReCANet produces consistently larger bad segments than TIFU-KNN across the post-processing approaches (15.2% compared to 10.0% for multiplicative scoring, 10.1% compared to 5.8% for weighted sum). This is consistent with the structural differences identified in earlier sections. ReCANet's candidate pool is restricted to previously purchased items, which limits its ability to replace displaced relevant items with alternative relevant items from collaborative signals. When a relevant item is pushed down by business-aware reranking, the replacement is drawn from the same narrow pool of personal history items, making it less likely to be relevant as well.

The model adjustments show the starkest contrast. The business-aware decay of TIFU-KNN produces a segment distribution with a moderate bad segment (14.3%) but a meaningful good segment (22.4%), with interpretable behavioral profiles that distinguish the two groups. The exponential loss weighting of ReCANet, by contrast, produces the worst segment distribution in the entire analysis (60.6% bad), with no systematic behavioral differentiation between segments. This result reinforces the conclusion of Section 5.5. The exp-weighted ReCANet does not post a viable business-aware integration method, as it degrades relevance for the majority of users without delivering compensating business value.

Practical implications. Unlike GT size, which is defined on the held-out test basket and therefore unobservable at inference time, basket size and diversity are stable customer characteristics directly computable from purchase history. This opens the possibility of targeted treatment in which the intensity of business-aware reranking is defined by observable customer traits, applying stronger reranking to customers whose historical purchasing patterns indicate robustness to it and lighter or no reranking to customers whose narrow, high-confidence purchasing profile makes them vulnerable to relevance loss. The empirical validation of such an approach, including the selection of behavioral thresholds and the quantification of the business value by exempting vulnerable customers, is left to future work.

6 Limitations & Future Work

6.1 Limitations

This study is designed as a proof-of-concept investigation that aims to characterise the feasibility and trade-off structure of business-aware grocery recommendation rather than to deliver a production-ready system. Several limitations follow directly from this scope and should be taken into account when interpreting the results.

The business value score is constructed from simulated data. The most significant limitation concerns the BVS dimensions as contribution margin, inventory pressure, temporal urgency, cross-selling potential, and strategic priority are populated using synthetically generated values derived from literature-documented parameter ranges rather than actual retailer data. While the simulation procedure is grounded in the operational realities of grocery retail and the resulting score distributions are consistent with known category-level differences, the specific BizValue figures reported throughout this thesis cannot be interpreted as true financial gains. A retailer deploying this framework would need to replace the simulated BVS with scores derived from its own margin accounting, inventory management, and merchandising systems. The appropriate calibration of such a score, including the selection of dimensions and their relative weighting, is itself an open research question. We therefore consider the results of this thesis as characterizing the shape of the trade-off landscape rather than quantifying its absolute magnitude.

The study relies on a single dataset from a single retail context. All experiments are conducted on the Dunnhumby dataset, which captures the transaction history of 2,500 households over a period of 711 days. While this is a well-regarded dataset for grocery recommendation research, it carries specific demographic, geographic, and assortment characteristics that may not generalise to other retail environments. The dataset exhibits a high repeat purchase rate of approximately 48%, which is particularly favourable to frequency-based modelling approaches such as TIFU-KNN. The comparative advantage of TIFU-KNN over ReCANet under business-aware optimisation may therefore in part reflect this property of the data rather than a universal architectural principle. Whether the structural findings reported here, in particular the relationship between candidate set diversity and optimisation responsiveness, hold across datasets with different purchase patterns, product mixes, or customer behaviours remains to be established.

Evaluation is conducted entirely offline. The experimental framework uses a leave-last-basket-out protocol, which measures whether recommended items overlap with what customers

actually purchased on their next shopping trip. This is the standard evaluation methodology in next-basket recommendation research and is appropriate for comparing methods under controlled conditions. However, offline evaluation cannot capture several dimensions that are critical to real-world deployment. It does not reveal whether customers would have accepted business-driven recommendations they would not have spontaneously chosen, nor does it measure the long-run effects of systematic business-aware reranking on customer trust, satisfaction, or retention. We note that the relationship between offline accuracy metrics and the business outcomes that ultimately matter to a retailer, such as revenue, margin, and customer lifetime value, is not established by this study. A meaningful assessment of real-world impact would require online experimentation.

The MIP-based methods face scalability constraints at full catalogue size. The epsilon-constraint and weighted sum formulations operate on a truncated candidate set of size C, pre-filtered by the base recommender model. This two-stage design is computationally tractable within the experimental setup and is representative of how such methods would typically be deployed. However, the experiments are conducted on a subset of 1,500 users rather than the full dataset population, and scaling MIP-based reranking to the full 92,339-item catalogue without pre-filtering would require substantially more efficient solver configurations or approximate optimisation strategies. The computational behaviour of these methods at full deployment scale is not characterised in this study.

The model comparison is limited to two architectures. The structural finding that candidate set diversity and score distribution shape determine a model's responsiveness to business-aware optimisation is derived from a comparison of TIFU-KNN and ReCANet. These two models represent meaningfully different design philosophies: neighbourhood-based versus neural, collaborative versus history-restricted. While this contrast is well-suited for the purpose of the study, the generality of the finding across a broader range of architectures, including transformer-based next-basket models and graph neural network approaches, has not been tested. Further research could investigate whether these relationships hold for other model families.

6.2 Future Work

The findings of this thesis open several directions for future research. These range from near-term extensions that directly address the limitations identified above to methodological advances that build on the framework developed here and longer-term applied directions necessary for real-world deployment.

BVS calibration with real retailer data. The most immediate extension is replacing the simulated business value score with one grounded in actual retailer data. This would involve partnering with a grocery retailer to obtain product-level margin data, live inventory records,

and supplier contract information, and using these to validate whether the five BVS dimensions and their relative weighting align with real operational priorities. Such a study would both test the practical relevance of the framework and address the most significant interpretive limitation of the current work. Beyond validation, it would surface questions about BVS design that the simulation cannot raise: how volatile business value scores are across days and seasons, how retailers should handle products for which reliable margin data is unavailable, and whether the five-dimensional structure remains appropriate across different store formats and assortment profiles.

Dynamic and real-time business value integration. The current framework treats the BVS as a static score computed once and applied uniformly across all recommendation events. In practice, however, the operational urgency of business objectives changes continuously. A product approaching its expiration date becomes progressively more important to recommend as the deadline approaches; inventory levels fluctuate intraday; promotional windows open and close. A natural extension is therefore to incorporate time-varying business signals into the recommendation pipeline, updating item-level business weights dynamically based on live inventory and pricing data. This would make the framework substantially more responsive to the perishability and stock management concerns identified as central to grocery retail in Chapter 2 and would represent a meaningful step toward a deployable system.

Extension to a three-stakeholder framework. This thesis models the recommendation problem as a two-sided trade-off between customer utility and retailer business value. Grocery retail, however, involves a third important stakeholder, the supplier. Suppliers routinely fund promotional activity in exchange for placement, and retailers increasingly manage category assortments in coordination with key supplier partners. Extending the BVS or the optimisation framework to incorporate supplier-level objectives, such as new product launch support, promotional co-funding commitments, or category development agreements, would move the framework toward a full three-sided market formulation. This direction connects naturally to the multistakeholder recommender systems literature reviewed in Chapter 2, where three-party optimisation remains relatively underexplored in empirical work.

Broader architectural comparison. The structural finding that candidate set diversity and score distribution shape determine a model's responsiveness to business-aware optimisation is derived from a comparison of two architectures. Testing this hypothesis across a wider range of models, including transformer-based next-basket approaches, graph neural networks, and hybrid collaborative-content architectures, would establish whether it constitutes a general principle or one specific to the TIFU-KNN versus ReCANet contrast. This is a relatively tractable empirical extension that would substantially strengthen the generalisability of one of the thesis's central contributions.

Adaptive trade-off control. All integration methods evaluated in this thesis require the practitioner to set a fixed trade-off parameter (λ for weighted sum, ε for the epsilon-constraint,

β for model adjustment), which is then applied uniformly across all users and recommendation events. A more sophisticated approach would learn to adapt this parameter dynamically based on contextual signals. The current inventory state, the time of day or week, the customer's purchase history and price sensitivity, or the retailer's current promotional calendar. Contextual bandit or reinforcement learning formulations could be used to optimise the trade-off parameter over time in response to observed outcomes, moving from a static configuration to one that continuously balances customer and business objectives based on real-world feedback.

Online evaluation and A/B testing. The most important gap that this thesis cannot close is the absence of online experimentation. Understanding how customers actually respond to business-aware recommendations, whether accuracy losses at higher business value settings translate into engagement drops, and whether the projected revenue gains are realised, requires a validation step that offline evaluation cannot provide. A live A/B test comparing a business-aware variant against the standard recommendation baseline, instrumented to track not only click-through and purchase rates but also longer-term indicators such as basket size and return frequency, would supply the evidence base needed before any deployment decision. We consider this the natural next step for any applied development building on the foundations established here.

7 Summary

This thesis shows that business objectives can be integrated into grocery recommender systems without undermining their customer-facing function, and it identifies the conditions under which such integration is most effective.

RQ1 asked how standard grocery recommender systems can be extended to incorporate business objectives without substantially compromising recommendation quality for customers. The evidence across all four integration methods supports a clear answer: meaningful gains in business value are attainable at a modest accuracy cost, and at conservative parameter settings the cost can be reduced to near zero.

The most efficient configuration observed was the business-aware decay extension to TIFU-KNN at $\beta = 0.3$, which produced a BizValue improvement of approximately 3% while keeping both Recall and NDCG within $\pm 0.3\%$ of the baseline. Post-processing methods produced larger BizValue gains but at proportionally larger accuracy costs. The trade-off is therefore not binary but continuous, and the practitioner can position the system anywhere along it depending on operational priorities.

RQ2 asked which integration approach, post-processing reranking or model-level adjustment, achieves a more favorable trade-off between customer utility and business value. The answer depends on the base model, but a consistent pattern emerges.

Post-processing reranking operates on a candidate set already fixed by the base model, so any business-oriented reordering necessarily displaces relevance-ranked items and incurs an accuracy cost from the moment a business signal is introduced. Integrating business awareness at the representation level produces a qualitatively different trade-off profile. For TIFU-KNN, the business-aware decay extension achieves a near free improvement that no post-processing method matches at comparable levels of business value gain. For ReCANet, exponential loss weighting yields smaller BizValue gains than post-processing but is the only method evaluated does not require trading aggregate recommendation quality for business value, although customer-level analysis reveals heterogeneous effects across users.

Representation-level integration, where the business signal is incorporated into the same structure from which relevance is derived, is therefore a more efficient point of intervention than post-hoc reranking on both architectures, although the magnitude of the advantage differs.

A complementary finding within the post-processing approaches is that the two MIP formulations should be viewed as complementary tools rather than competing alternatives. The weighted-sum method produces a nearly linear Pareto frontier and behaves predictably across parameter settings, which makes it well suited to stable day-to-day optimization. The epsilon-constraint method produces a concave frontier and guarantees a minimum business value floor regardless of the candidate score distribution, which makes it the natural choice for hard operational commitments such as inventory liquidation targets or supplier agreements. Nei-

ther formulation dominates the other. The choice depends on whether the business objective is best expressed as a weighting preference or as a hard floor.

RQ3 emerged from the empirical results and asked how the structural properties of a recommender system influence its responsiveness to business-aware optimization. Across all four integration methods, TIFU-KNN consistently responded more favorably than ReCANet.

Three properties of the base model account for the difference: the diversity of its candidate set, the variance of business value scores within that set, and the shape of its prediction score distribution. Models that produce diverse candidates with relatively flat score distributions, such as the collaborative filtering component of TIFU-KNN, give the optimizer room to substitute high-value items for marginally less relevant alternatives. Models that restrict candidates to the personal purchase history and produce sharply peaked score distributions, such as ReCANet, impose an upper bound on the achievable business value that no post-processing method can exceed.

This structural relationship appears to have received limited attention in the multistakeholder recommender systems literature, where focus has predominantly been on integration methods rather than on the architectural prerequisites that make those methods effective. The implication is that the design problem for business-aware grocery recommendation is not only which integration method to apply, but also which base model produces a candidate set and score distribution that are amenable to integration in the first place, and at which stage of the pipeline the business signal can be introduced most efficiently.

Beyond the three research questions, the customer-level analysis surfaces a finding that cuts across all integration methods: business-aware optimization does not affect all users equally, and the affected populations differ systematically between integration paradigms. Post-processing methods harm users whose held-out baskets contain few relevant items, an effect rooted in the structure of the evaluation instance rather than in stable customer behavior. Model-level adjustments harm a different population, specifically habitual shoppers with rigid, low-diversity purchase routines, whose established patterns are most strongly encoded in the user representation and therefore most disrupted by its reshaping. Since observable behavioral traits such as average basket size and item diversity are stable and computable from the purchase history, this finding opens a practical path toward heterogeneous treatment policies that adjust the intensity of business-aware reranking based on individual customer profiles.

Several limitations should be acknowledged. The Business Value Score relies on simulated data rather than real operational metrics, and its calibration for a specific retailer remains an open question. The Dunnhumby dataset, while widely used, represents a single market and time period. The MIP-based approaches, while producing provably optimal solutions within their formulation, face computational scaling challenges at full catalog sizes. This thesis is therefore positioned as a proof-of-concept study that characterizes the trade-off landscape rather than a production-ready system. The framework, the structural insight regarding architectural responsiveness, and the customer-level impact patterns collectively provide a principled foundation for subsequent applied development and online validation.

8 References

References

Abdollahpouri, H. & Burke, R. (2022). Multistakeholder recommender systems. In F. Ricci, L. Rokach & B. Shapira (Hrsg.), *Recommender systems handbook* (S. 647–677). Springer.

Abdollahpouri, H., Burke, R. & Mobasher, B. (2019). Managing popularity bias in recommender systems with personalized re-ranking. In *Proceedings of the thirty-second international florida artificial intelligence research society conference* (S. 413–418).

Abdollahpouri, H. et al. (2020). Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction*, *30* (1), 127–158.

Adomavicius, G. & Kwon, Y. (2012). Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering*, *24* (5), 896–911.

Adomavicius, G. & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17* (6), 734–749.

Ailawadi, K. L. & Farris, P. W. (2017). Managing multi- and omni-channel distribution: Metrics and research directions. *Journal of Retailing*, *93* (1), 120–135.

Amorim, P., Meyr, H., Almeder, C. & Almada-Lobo, B. (2013). Managing perishability in production-distribution planning: A discussion and review. *Flexible Services and Manufacturing Journal*, *25* (3), 389–413.

Anderson, A., Kumar, R., Tomkins, A. & Vassilvitskii, S. (2014). The dynamics of repeat consumption. In *Proceedings of the 23rd international conference on world wide web* (S. 419–430).

Ariannezhad, M. et al. (2022). ReCANet: A repeat consumption-aware neural network for next basket recommendation in grocery shopping. In *Proceedings of the 45th international acm sigir conference on research and development in information retrieval* (S. 1240–1250).

Bhagat, S., Muralidharan, K., Lobzhanidze, A. & Vishwanath, S. (2018). Buy it again: Modeling repeat purchase recommendations. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (S. 62–70).

Brijs, T., Swinnen, G., Vanhoof, K. & Wets, G. (1999). Using association rules for product assortment decisions: A case study. In *Proceedings of the fifth acm sigkdd international conference on knowledge discovery and data mining* (S. 254–260).

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, *12* (4), 331–370.

Christodoulou, K. et al. (2017). A real-time targeted recommender system. In *Proceedings of the 19th international conference on enterprise information systems* (S. 703–712).

Dhar, S. K., Morrison, D. G. & Raju, J. S. (2001). The effect of package coupons on brand choice: An epilogue on profits. *Marketing Science*, *20* (1), 19–28.

Ekstrand, M. D., Burke, R. & Diaz, F. (2019). Fairness and discrimination in recommendation and retrieval. In *Proceedings of the 12th acm conference on recommender systems* (S. 1–2). ACM.

Guidotti, R. et al. (2017). Market basket prediction using user-centric temporal annotated recurring sequences. In *Proceedings of the 2017 ieee international conference on data mining (icdm)* (S. 895–900).

Guidotti, R. et al. (2019). Personalized market basket prediction with temporal annotated recurring sequences. *IEEE Transactions on Knowledge and Data Engineering*, *31* (11), 2151–2163.

He, X. et al. (2017). Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web* (S. 173–182). International World Wide Web Conferences Steering Committee.

Herbon, A., Levner, E. & Cheng, T. (2014). Perishable inventory management with dynamic pricing using time–temperature indicators linked to automatic detecting devices. *International Journal of Production Economics*, *147*, 605–613.

Hu, H., He, X., Gao, J. & Zhang, Z.-L. (2020). Modeling personalized item frequency information for next-basket recommendation. In *Proceedings of the 43rd international acm sigir conference on research and development in information retrieval* (S. 1071–1080). ACM. doi: 10.1145/3397271.3401066

Jannach, D. & Abdollahpouri, H. (2023). A survey on multi-objective recommender systems. *Frontiers in Big Data*, *6*, 1157899.

Jannach, D. & Adomavicius, G. (2017). Price and profit awareness in recommender systems. *arXiv preprint arXiv:1707.08029*.

Jannach, D. & Bauer, C. (2020). Escaping the McNamara fallacy: Toward more impactful recommender systems research. *AI Magazine*, *41* (4), 79–95.

Jansen, L. Z. H., Bennin, K. E., van Kleef, E. & Van Loo, E. J. (2024). Online grocery shopping recommender systems: Common approaches and practices. *Computers in Human Behavior*, *159*, 108336.

Jugovac, M., Jannach, D. & Lerche, L. (2017). Efficient optimization of multiple recommendation quality factors according to individual user tendencies. *Expert Systems with Applications*, *81*, 321–331.

Koren, Y. (2008). Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th acm sigkdd international conference on knowledge discovery and data mining* (S. 426–434). ACM.

Krafft, M. & Mantrala, M. K. (Hrsg.). (2010). *Retailing in the 21st century: Current and future trends* (2. Aufl.). Springer.

Kutuzova, T. & Melnik, M. (2018). Market basket analysis of heterogeneous data sources for recommendation system improvement. *Procedia Computer Science*, *136*, 246–254.

Lau, H. C., Nakandala, D. & Shum, P. K. (2016). A business process decision model for fresh-food supplier evaluation. *Business Process Management Journal*, *22* (1), 148–169.

Le, D.-T., Lauw, H. W. & Fang, Y. (2019). Correlation-sensitive next-basket recommendation. In *Proceedings of the 28th international joint conference on artificial intelligence* (S. 2808–

2814).

Lee, A., Lan, J. C.-W., Jambrak, A. R. et al. (2024). Upcycling fruit waste into microalgae biotechnology: Perspective views and way forward. *Food Chemistry: Molecular Sciences*, *8*, 100203. doi: 10.1016/j.fochms.2024.100203

Li, M., Dias, B. M., Jarman, I., El-Deredy, W. & Lisboa, P. J. (2009). Grocery shopping recommendations based on basket-sensitive random walk. In *Proceedings of the 15th acm sigkdd international conference on knowledge discovery and data mining* (S. 1215–1224).

Liang, D., Krishnan, R. G., Hoffman, M. D. & Jebara, T. (2018). Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 world wide web conference* (S. 689–698). International World Wide Web Conferences Steering Committee.

Malthouse, E. C. et al. (2019). A multistakeholder recommender systems algorithm for allocating sponsored recommendations. In *Proceedings of the workshop on recommendation in multistakeholder environments (rmse) at recsys 2019.*

Nguyen, T.-T. et al. (2017). Exploring the filter bubble: The effect of using recommender systems on content diversity. In *Proceedings of the 26th international conference on world wide web* (S. 677–686).

Park, Y.-J. et al. (2018). Group recommender system for store product placement. *Data Mining and Knowledge Discovery*, *32* (6), 1952–1986.

Rijpkema, W. A., Rossi, R. & van der Vorst, J. G. (2014). Effective sourcing strategies for perishable product supply chains. *International Journal of Physical Distribution & Logistics Management*, *44* (6), 494–510.

Rodriguez, M., Posse, C. & Zhang, E. (2012). Multiple objective optimization in recommender systems. In *Proceedings of the sixth acm conference on recommender systems* (S. 11–18). ACM.

Roy, D. & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, *9* (1), 59.

Sano, N., Machino, N., Yada, K. & Suzuki, T. (2015). Recommendation system for grocery store considering data sparsity. *Procedia Computer Science*, *60*, 1406–1413.

Sürer, Ö., Burke, R. & Malthouse, E. C. (2018). Multistakeholder recommendation with provider constraints. In *Proceedings of the 12th acm conference on recommender systems* (S. 54–62). ACM.

Tabane, T. L., Phume, T. B. & Retief, M.-M. (2024). Effects of physical stock loss on the financial performance of retail enterprises. *South African Journal of Economic and Management Sciences*, *27*. doi: 10.4102/sajems.v27i1.5410

Todd, E. C. D. & Faour-Klingbeil, D. (2024). Impact of food waste on society, specifically at retail and foodservice levels in developed and developing countries. *Foods*, *13*, 2098. doi: 10.3390/foods13132098

Trejo-Pech, C. J. O. & White, S. (2024). Financial ratios of the u.s. grocery sector in a changing industry landscape. *Applied Economics Teaching Resources (AETR)*, *6* (2). doi: 10.22004/ag.econ.343483

Zhang, S., Yao, L., Sun, A. & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, *52* (1), 1–38.

Zheng, Y. et al. (2021). A survey of recommender systems with multi-objective optimization. *Neurocomputing*, *474*, 141–153.

9 Appendix

9.1 Business Value Range Assignment

Table 9.1: Value Assignments and Distribution Ranges per Framework Category

Category	Category Level	Assigned Range / Distribution Logic
Contribution Margin (CM)	Low Medium High Very High	5% – 15% (0.05 – 0.15) 15% – 30% (0.15 – 0.30) 30% – 50% (0.30 – 0.50) 50% – 70% (0.50 – 0.70)
Inventory Pressure (IP)	Very Low Low Medium High Very High	1 – 10 units 10 – 50 units 50 – 200 units 200 – 500 units 500 – 1000 units
Temporal Urgency (TU)	Very Short Short Medium Long Very Long	1 – 7 days 1 – 30 days 1 – 180 days 1 – 730 days 1 – 3650 days
Cross-Selling Potential (CSP)	Low Medium High Very High	0 – 5 6 – 10 11 – 15 15 – 20
Strategic Priority (SP)	N/A	Binary logic: 5% probability of 1 and 95% probability of 0.

Note: All numeric metrics were subsequently passed through a MinMaxScaler to standardize them on a strict 0 to 20 scale prior to calculating the final aggregate score.

9.2 Commodity Category Assignment

Table 9.2: Commodity Category Mappings

Commodity	CM	IP	TU	CSP	Commodity	CM	IP	TU	CSP
FRZN ICE	H	M	VL	M	CORN	M	L	S	H
NO COMMODITY DESCRIPTION	M	L	L	M	CHEESES	M	M	M	H
BREAD	L	VH	S	VH	MISCELLANEOUS	H	L	L	M
FRUIT - SHELF STABLE	M	H	L	H	CITRUS	M	L	S	H
COOKIES/CONES	M	M	L	H	FROZEN	M	M	L	H
SPICES & EXTRACTS	M	M	VL	H	DISHWASH DETERGENTS	H	M	VL	M
VITAMINS	VH	M	L	H	SWEET GOODS & SNACKS	H	H	L	H
BREAKFAST SWEETS	M	H	M	M	NUTS	H	H	L	M
PNT BTR/JELLY/JAMS	M	H	L	H	TROPICAL FRUIT	M	L	VS	H
ICE CREAM/MILK/SHERBTS	M	H	L	H	CHICKEN/POULTRY	M	M	VS	VH
MAGAZINE	H	L	M	L	BROCCOLI/CAULIFLOWER	L	L	VS	L
AIR CARE	H	M	VL	L	SINUS AND ALLERGY	VH	M	L	M
CHEESE	M	H	M	H	MUSHROOMS	M	L	VS	M
SHORTENING/OIL	M	M	L	H	CANNED MILK	H	H	VL	H
COFFEE	M	VH	L	H	ONIONS	M	L	M	VH
DIETARY AID PRODUCTS	VH	M	L	M	INFANT CARE PRODUCTS	VH	M	L	M
PAPER HOUSEWARES	H	H	VL	H	TURKEY	M	M	VS	VH
BAKED BREAD/BUNS/ROLLS	M	M	S	VH	WATER	L	VH	VL	VH
VEGETABLES - SHELF STABLE	M	H	L	H	CAT LITTER	H	M	VL	M
HISPANIC	H	M	M	M	SEAFOOD-FRESH	M	L	VS	H
DINNER MXS:DRY	M	M	L	M	PIES	M	M	S	M
CONDIMENTS/SAUCES	M	H	L	H	TOBACCO OTHER	L	L	L	L
FRZN VEGETABLE/VEG DSH	M	M	L	H	SEAFOOD - MISC	M	M	S	H
BAKING NEEDS	M	M	L	H	FILM AND CAMERA PRODUCTS	H	VL	VL	M
DINNER SAUSAGE	M	M	M	H	PEPPERS-ALL	M	L	S	M
FRZN FRUITS	H	M	L	H	PLASTIC HOUSEWARES	H	M	VL	L
SEAFOOD - FROZEN	M	M	L	H	FOOT CARE PRODUCTS	VH	M	L	M
HOUSEHOLD CLEANG NEEDS	H	M	VL	M	SHOE CARE	H	M	VL	L
FD WRAPS/BAGS/TRSH BG	H	M	VL	H	FIREWORKS	VH	L	VL	L
DRY MIX DESSERTS	M	M	L	M	FLORAL-FLOWERING PLANTS	VH	L	VS	L
PICKLE/RELISH/PKLD VEG	M	M	L	H	SUNTAN	VH	M	L	M
CAKES	M	M	S	M	CEREAL/BREAKFAST	M	H	L	H
BAKING MIXES	M	M	L	H	CANDLES/ACCESSORIES	VH	L	VL	L
POTATOES	M	H	M	VH	COOKWARE & BAKEWARE	H	L	VL	L
FLUID MILK PRODUCTS	L	VH	VS	VH	DISPOSIBLE FOILWARE	H	M	VL	H
SOUP	M	H	L	H	FLORAL-FRESH CUT	VH	L	VS	L
BAKED SWEET GOODS	M	M	S	M	VEGETABLES SALAD	M	M	VS	H
COOKIES	M	H	L	H	AUDIO/VIDEO PRODUCTS	H	VL	VL	L
DRY BN/VEG/POTATO/RICE	M	H	L	VH	REFRIGERATED	M	M	M	M
FACIAL TISS/DNR NAPKIN	H	VH	VL	M	FROZEN CHICKEN	M	M	L	VH
FROZEN PIZZA	M	H	L	H	SMOKED MEATS	M	M	M	VH
EGGS	L	VH	S	VH	VALUE ADDED VEGETABLES	H	M	S	H
REFRGRATD DOUGH PRODUCTS	M	M	M	M	EYE AND EAR CARE PRODUCTS	VH	M	L	M
HOT CEREAL	M	H	L	H	SNKS/CKYS/CRKR/CNDY	M	H	M	M
COLD CEREAL	M	VH	L	H	FLORAL BALLOONS	VH	VL	S	L
SUGARS/SWEETNERS	M	VH	VL	H	AUTOMOTIVE PRODUCTS	H	M	VL	L
SEAFOOD - SHELF STABLE	M	M	L	H	LAXATIVES	VH	M	L	M
POPCORN	M	H	L	H	DOMESTIC WINE	L	L	VL	H
CANNED JUICES	M	H	L	H	MISC WINE	L	L	VL	H
STATIONERY & SCHOOL SUPPLIES	H	L	VL	M	OVERNIGHT PHOTOFINISHING	H	VL	VS	M
COLD AND FLU	VH	M	L	M	TOMATOES	M	L	VS	H
BABY HBC	VH	M	L	M	PEARS	M	L	S	M
BAG SNACKS	M	H	M	H	FITNESS&DIET	H	L	L	M
BEANS - CANNED GLASS & MW	H	H	L	H	BATH	VH	M	L	M

(continued on next page)

(continued from previous page)

Commodity	CM	IP	TU	CSP	Commodity	CM	IP	TU	CSP
FROZEN MEAT	M	M	L	H	COUPON/MISC ITEMS	L	VL	VS	L
SOAP - LIQUID & BAR	H	M	VL	H	PORK	M	M	VS	VH
ROLLS	M	M	S	M	PERSONAL CARE APPLIANCES	VH	VL	VL	M
NON-DAIRY BEVERAGES	M	H	M	M	CONDIMENTS	M	H	L	H
KITCHEN GADGETS	H	L	VL	L	BEVERAGE	M	H	M	M
CRACKERS/MISC BKD FD	M	H	M	H	SNACKS	M	H	M	H
CONVENIENT BRKFST/WHLSM SNACKS	M	M	M	H	VALUE ADDED FRUIT	H	L	VS	H
CHIPS&SNACKS	M	H	M	H	SALAD BAR	M	L	VS	H
CANDY - PACKAGED	M	H	L	H	HAIR CARE ACCESSORIES	VH	M	VL	L
SOFT DRINKS	M	VH	L	H	DIAPERS & DISPOSABLES	VH	M	L	VH
HAIR CARE PRODUCTS	VH	M	L	L	SMOKING CESSATIONS	VH	VL	L	M
CANDY - CHECKLANE	M	H	L	H	FUEL	L	L	VL	M
BUTTER	L	H	M	H	PREPARED FOOD	VH	L	VS	M
FRZN MEAT/MEAT DINNERS	M	M	L	VH	PET CARE SUPPLIES	H	M	L	VH
SHAVING CARE PRODUCTS	VH	M	VL	M	COFFEE SHOP	VH	L	VS	H
WATER - CARBONATED/FLVRD DRINK	M	H	L	VH	COUPON	L	VL	VS	L
FRZN BREAKFAST FOODS	M	M	L	M	JCAL	L	VL	VS	M
MILK BY-PRODUCTS	M	M	S	M	ROSES	VH	L	VS	L
FIRST AID PRODUCTS	VH	M	L	M	PROD SUPPLIES	H	L	L	M
NEWSPAPER	L	L	VS	H	J-HOOKS	H	VL	VL	M
SALADS/DIPS	M	M	VS	H	RICE CAKES	H	M	L	H
INSECTICIDES	VH	M	VL	L	LAMB	M	L	VS	M
ORGANICS FRUIT & VEGETABLES	VH	L	VS	H	IMPORTED WINE	L	L	VL	H
ELECTRICAL SUPPPLIES	H	L	VL	M	UNKNOWN	M	VL	L	M
LAUNDRY DETERGENTS	H	VH	VL	VH	DELI SPECIALTIES (RETAIL PK)	VH	L	M	H
JUICE	M	VH	L	H	PREPARED/PKGD FOODS	M	M	M	M
ISOTONIC DRINKS	M	M	L	M	FAMILY PLANNING	VH	M	L	M
FRZN JCE CONC/DRNKS	M	H	L	M	EASTER	VH	L	S	L
LAUNDRY ADDITIVES	H	M	VL	M	CIGARS	L	L	L	L
IRONING AND CHEMICALS	H	M	VL	M	PACKAGED NATURAL SNACKS	M	M	M	H
TEAS	M	M	L	H	FLORAL-FOLIAGE PLANTS	VH	L	M	L
DRY NOODLES/PASTA	M	H	VL	H	MISCELLANEOUS CROUTONS	M	M	L	M
PASTA SAUCE	M	H	L	H	COUPONS/STORE & MFG	L	VL	VS	L
PROCESSED	L	M	L	M	APPAREL	VH	L	VL	L
MAKEUP AND TREATMENT	VH	L	L	M	PREPAID WIRELESS&ACCESSORIES	VH	VL	VL	M
ANALGESICS	VH	M	L	M	HOME FREEZING & CANNING SUPPLY	H	M	VL	M
HOSIERY/SOCKS	VH	L	VL	L	IN-STORE PHOTOFINISHING	VH	VL	VS	M
CAT FOOD	H	L	L	VH	ADULT INCONTINENCE	VH	M	L	M
BATTERIES	H	L	VL	M	PARTY TRAYS	M	L	VS	L
MOLASSES/SYRUP/PANCAKE MIXS	M	H	L	H	DOMESTIC GOODS	H	L	VL	M
BOOKSTORE	H	L	VL	L	FALL AND WINTER SEASONAL	VH	M	L	L
BATH TISSUES	H	VH	VL	H	TOYS AND GAMES	VH	L	VL	L
FROZEN PIE/DESSERTS	M	M	L	H	LAWN AND GARDEN SHOP	VH	L	VL	M
MEAT - SHELF STABLE	M	M	L	M	GLASSWARE & DINNERWARE	H	L	VL	M
MEAT - MISC	M	M	S	VH	BAKING	M	H	L	H
SPRING/SUMMER SEASONAL	VH	M	L	L	RESTRICTED DIET	VH	L	L	M
LUNCHMEAT	M	H	VS	VH	FRAGRANCES	VH	VL	VL	M
BREAKFAST SAUSAGE/SANDWICHES	M	M	M	H	HALLOWEEN	VH	L	S	L
DRIED FRUIT	M	M	L	H	BAKERY PARTY TRAYS	M	L	VS	L
CHARCOAL AND LIGHTER FLUID	M	M	VL	L	SEWING	H	VL	VL	M
SALD DRSNG/SNDWCH SPRD	M	M	M	M	NATURAL HBC	VH	L	L	M
LIQUOR	L	L	VL	H	BIRD SEED	H	L	M	M
FROZEN BREAD/DOUGH	M	M	L	VH	GARDEN CENTER	VH	VL	L	M
HAND/BODY/FACIAL PRODUCTS	VH	M	L	M	COSMETIC ACCESSORIES	VH	M	VL	M
SNACK NUTS	M	H	M	H	ETHNIC PERSONAL CARE	VH	M	L	M
ORAL HYGIENE PRODUCTS	VH	M	VL	M	GLASSES/VISION AIDS	VH	VL	VL	M
BEERS/ALES	L	L	L	H	FLORAL- HARD GOODS	VH	L	VL	L
INFANT FORMULA	VH	L	M	VH	BABYFOOD	VH	M	L	M
REFRGRATD JUICES/DRNKS	M	H	M	H	TOYS	VH	L	VL	L

(continued on next page)

(continued from previous page)

Commodity	CM	IP	TU	CSP	Commodity	CM	IP	TU	CSP
YOGURT	M	M	S	H	CHRISTMAS SEASONAL	VH	M	L	L
DEODORANTS	VH	M	VL	H	COFFEE SHOP SWEET GOODS&RETAIL	VH	L	VS	H
BACON	M	H	VS	H	DRY TEA/COFFEE/COCO MIX	M	H	L	H
DOG FOODS	H	L	L	VH	VEAL	M	L	VS	M
FRZN NOVELTIES/WTR ICE	M	M	L	M	CONTINUITIES	M	M	M	M
FEMININE HYGIENE	VH	M	VL	H	PORTABLE ELECTRIC APPLIANCES	VH	VL	VL	M
COFFEE FILTERS	H	M	L	H	VALENTINE	VH	L	VS	L
BROOMS AND MOPS	H	M	VL	M	FROZEN - BOXED(GROCERY)	M	M	L	H
GREETING CARDS/WRAP/PARTY SPLY	VH	L	VL	H	SERVICE BEVERAGE	M	H	VS	L
WAREHOUSE SNACKS	M	H	M	H	GIFT & FRUIT BASKETS	VH	VL	VS	H
DRY SAUCES/GRAVY	M	M	L	H	SUSHI	VH	L	VS	H
MARGARINES	M	M	M	H	PROPANE	L	L	VL	M
PWDR/CRYSTL DRNK MX	M	M	L	M	FRZN SEAFOOD	M	M	L	H
OLIVES	M	M	L	H	HOME FURNISHINGS	H	VL	VL	M
MISC. DAIRY	M	M	M	M	PHARMACY	VH	L	L	H
COCOA MIXES	M	M	L	M	NATURAL VITAMINS	VH	L	L	H
SALAD MIX	M	M	VS	H	DELI SUPPLIES	H	M	VS	H
HARDWARE SUPPLIES	H	L	VL	L	WATCHES/CALCULATORS/LOBBY	VH	VL	VL	M
HOT DOGS	M	H	VS	M	SEASONAL	VH	L	L	L
FLOUR & MEALS	M	H	VL	H	BOUQUET (NON ROSE)	VH	L	VS	L
SYRUPS/TOPPINGS	M	M	L	H	HOME HEALTH CARE	VH	VL	L	M
ANTACIDS	VH	M	L	M	RW FRESH PROCESSED MEAT	M	M	VS	VH
BLEACH	H	M	L	M	EXOTIC GAME/FOWL	M	VL	VS	L
PAPER TOWELS	H	VH	VL	H	NDAIRY/TEAS/JUICE/SOD	M	M	M	H
STONE FRUIT	M	L	VS	H	TICKETS	L	VL	VL	L
MELONS	M	L	VS	M	MEAT SUPPLIES	H	L	S	VH
BERRIES	M	L	VS	H	NEW AGE	M	M	M	M
CHICKEN	M	M	VS	VH	FLORAL-ACCESSORIES	VH	L	VL	L
SANDWICHES	M	M	VS	M	EASTER LILY	VH	L	VS	L
GRAPES	M	L	VS	H	MISCELLANEOUS HBC	VH	L	L	M
CARROTS	M	L	VS	M	BULK FOODS	M	M	L	M
APPLES	L	L	S	H	NON EDIBLE PRODUCTS	H	H	L	M
HERBS	M	L	VS	M	QUICK SERVICE	M	L	VS	L
CIGARETTES	L	VH	VL	L	BOTTLE DEPOSITS	L	VH	VL	L
HEAT/SERVE	M	M	M	M	JAPANESE	L	L	L	L
BABY FOODS	VH	M	L	VH	DOLLAR VALUE PRODUCTS	M	M	L	M
SQUASH	H	L	S	M	MISCELLANEOUS	L	VL	VS	M
VEGETABLES - ALL OTHERS	M	M	S	H	SPORTS MEMORABLILIA	H	VL	VL	M
DELI MEATS	M	H	VS	VH	LONG DISTANCE CALLING CARDS	VH	L	VL	M
BEEF	M	M	VS	VH	PKG.SEAFOOD MISC	M	M	S	H
FRZN POTATOES	M	M	L	VH	FROZEN PACKAGE MEAT	M	M	L	H

Note: CM = Contribution Margin, IP = Inventory Pressure, TU = Temporal Urgency, CSP = Cross-Selling Potential. Levels — VL: Very Low / Very Long (TU); L: Low / Long; M: Medium; H: High; VH: Very High; VS: Very Short (TU); S: Short (TU).

Table 9.3: Performance comparison of reranking approaches for **TIFU-KNN**. Orig = original score; Mult $\Delta\%$ = multiplicative baseline; WS $\Delta\%$ (α) = weighted-sum MIP per α; ε $\Delta\%$ (ε_p) = epsilon-constraint MIP per percentile; Adj. TIFU $\Delta\%$ (β) = adjusted TIFU per β.

	Recall					NDCG					Biz Value				
K	**Orig**	**Mult $\Delta\%$**	**WS $\Delta\%$ (α)**	**ε $\Delta\%$ (ε_p)**	**Adj. TIFU $\Delta\%$ (β)**	**Orig**	**Mult $\Delta\%$**	**WS $\Delta\%$ (α)**	**ε $\Delta\%$ (ε_p)**	**Adj. TIFU $\Delta\%$ (β)**	**Orig**	**Mult $\Delta\%$**	**WS $\Delta\%$ (α)**	**ε $\Delta\%$ (ε_p)**	**Adj. TIFU $\Delta\%$ (β)**
5	0.1084	−7.76%	0.1 −61.47% 0.2 −52.35% 0.3 −42.03% 0.4 −25.88% 0.5 −13.64% 0.6 −8.91% 0.7 −5.50% 0.8 −3.55% 0.9 −1.02%	90 −24.56% 80 −13.91% 70 −6.76% 60 −3.17% 50 −3.35% 40 −2.29% 30 −1.45% 20 +0.19% 10 +0.39%	0.1 −0.51% 0.2 −0.75% 0.3 −1.62% 0.4 −3.70% 0.5 −4.31% 0.6 −4.50% 0.7 −6.49% 0.8 −7.94% 0.9 −8.25%	0.1885	−0.93%	0.1 −53.42% 0.2 −39.77% 0.3 −29.31% 0.4 −15.79% 0.5 −5.88% 0.6 −1.81% 0.7 +0.52% 0.8 +1.00% 0.9 +1.08%	90 −11.00% 80 −4.68% 70 −1.94% 60 −1.03% 50 −1.05% 40 −0.44% 30 −0.31% 20 +0.22% 10 +0.16%	0.1 +0.11% 0.2 +0.23% 0.3 +0.24% 0.4 −0.05% 0.5 −0.39% 0.6 −0.47% 0.7 −1.31% 0.8 −1.88% 0.9 −2.33%	177.46	+10.89%	0.1 +30.88% 0.2 +29.45% 0.3 +26.85% 0.4 +22.21% 0.5 +17.04% 0.6 +12.45% 0.7 +9.12% 0.8 +6.31% 0.9 +3.75%	90 +19.87% 80 +12.40% 70 +7.27% 60 +3.86% 50 +2.11% 40 +0.92% 30 +0.35% 20 +0.01% 10 −0.22%	0.1 +1.73% 0.2 +2.58% 0.3 +3.50% 0.4 +4.47% 0.5 +5.42% 0.6 +6.41% 0.7 +7.43% 0.8 +8.64% 0.9 +9.75%
10	0.1513	−3.68%	0.1 −57.18% 0.2 −50.53% 0.3 −39.51% 0.4 −25.99% 0.5 −11.95% 0.6 −5.82% 0.7 −2.23% 0.8 −1.71% 0.9 +1.20%	90 −24.06% 80 −9.56% 70 −3.97% 60 −0.44% 50 +1.38% 40 +1.50% 30 +1.37% 20 +1.41% 10 +1.41%	0.1 −0.60% 0.2 −0.37% 0.3 −0.36% 0.4 −1.41% 0.5 −1.49% 0.6 −3.52% 0.7 −4.05% 0.8 −4.53% 0.9 −5.76%	0.1802	−1.90%	0.1 −52.75% 0.2 −41.91% 0.3 −32.11% 0.4 −19.22% 0.5 −7.79% 0.6 −2.69% 0.7 −0.58% 0.8 +0.39% 0.9 +1.08%	90 −13.36% 80 −3.79% 70 −1.66% 60 −0.29% 50 +0.23% 40 +0.43% 30 +0.41% 20 +0.45% 10 +0.45%	0.1 −0.50% 0.2 −0.32% 0.3 −0.31% 0.4 −0.86% 0.5 −0.85% 0.6 −1.59% 0.7 −2.09% 0.8 −2.60% 0.9 −3.59%	349.77	+10.29%	0.1 +28.55% 0.2 +27.68% 0.3 +25.60% 0.4 +22.01% 0.5 +17.97% 0.6 +13.86% 0.7 +10.27% 0.8 +7.10% 0.9 +4.25%	90 +20.92% 80 +12.71% 70 +6.80% 60 +3.06% 50 +1.16% 40 +0.30% 30 +0.03% 20 −0.14% 10 −0.23%	0.1 +1.82% 0.2 +2.50% 0.3 +3.26% 0.4 +4.19% 0.5 +5.11% 0.6 +6.09% 0.7 +7.01% 0.8 +8.09% 0.9 +9.17%
20	0.1954	−2.92%	0.1 −48.69% 0.2 −43.30% 0.3 −32.59% 0.4 −19.56% 0.5 −11.89% 0.6 −6.98% 0.7 −2.55% 0.8 −0.74% 0.9 +1.49%	90 −34.65% 80 −8.01% 70 −1.38% 60 +1.72% 50 +0.73% 40 +0.30% 30 +0.15% 20 +0.15% 10 +0.15%	0.1 −0.09% 0.2 −0.53% 0.3 +0.06% 0.4 +0.34% 0.5 −0.88% 0.6 −1.34% 0.7 −1.21% 0.8 −2.68% 0.9 −3.99%	0.1783	−1.75%	0.1 −49.20% 0.2 −39.81% 0.3 −30.30% 0.4 −18.03% 0.5 −8.57% 0.6 −3.97% 0.7 −0.84% 0.8 +0.54% 0.9 +1.17%	90 −22.94% 80 −3.54% 70 −0.13% 60 +0.93% 50 +0.38% 40 +0.21% 30 +0.15% 20 +0.15% 10 +0.15%	0.1 −0.26% 0.2 −0.45% 0.3 −0.42% 0.4 −0.58% 0.5 −0.93% 0.6 −1.22% 0.7 −1.52% 0.8 −2.41% 0.9 −3.29%	687.68	+9.49%	0.1 +24.57% 0.2 +23.95% 0.3 +22.46% 0.4 +20.17% 0.5 +17.35% 0.6 +14.25% 0.7 +10.79% 0.8 +7.59% 0.9 +4.60%	90 +22.24% 80 +14.09% 70 +7.38% 60 +2.86% 50 +0.61% 40 −0.12% 30 −0.31% 20 −0.42% 10 −0.51%	0.1 +1.82% 0.2 +2.27% 0.3 +2.88% 0.4 +3.63% 0.5 +4.39% 0.6 +5.17% 0.7 +6.05% 0.8 +6.87% 0.9 +7.79%
50	0.2573	−1.70%	0.1 −30.82% 0.2 −22.79% 0.3 −14.68% 0.4 −10.58% 0.5 −6.77% 0.6 −3.72% 0.7 −2.47% 0.8 −1.26% 0.9 −0.43%	90 −17.78% 80 −21.44% 70 −2.70% 60 −0.24% 50 +0.39% 40 +0.23% 30 +0.23% 20 +0.26% 10 +0.26%	0.1 −0.26% 0.2 +0.28% 0.3 +0.03% 0.4 +0.32% 0.5 +0.57% 0.6 +0.48% 0.7 +0.16% 0.8 −0.20% 0.9 +0.09%	0.1923	−1.95%	0.1 −41.42% 0.2 −32.18% 0.3 −24.14% 0.4 −14.87% 0.5 −7.32% 0.6 −3.49% 0.7 −1.50% 0.8 −0.20% 0.9 +0.21%	90 −13.83% 80 −17.26% 70 −1.06% 60 −0.02% 50 +0.16% 40 +0.09% 30 +0.09% 20 +0.10% 10 +0.10%	0.1 −0.29% 0.2 −0.18% 0.3 −0.43% 0.4 −0.64% 0.5 −0.59% 0.6 −0.87% 0.7 −1.27% 0.8 −1.81% 0.9 −2.32%	1690.62	+8.27%	0.1 +14.44% 0.2 +14.15% 0.3 +13.66% 0.4 +12.94% 0.5 +11.93% 0.6 +10.60% 0.7 +8.82% 0.8 +6.71% 0.9 +4.28%	90 +5.19% 80 +8.34% 70 +9.02% 60 +3.26% 50 +0.08% 40 −0.60% 30 −0.75% 20 −0.83% 10 −0.84%	0.1 +1.06% 0.2 +1.34% 0.3 +1.77% 0.4 +2.23% 0.5 +2.71% 0.6 +3.23% 0.7 +3.80% 0.8 +4.42% 0.9 +5.09%

Table 9.4: Performance comparison of reranking approaches for **ReCANet**. Orig = original score; Mult $\Delta\%$ = multiplicative baseline; WS $\Delta\%$ (α) = weighted-sum MIP per α; ε $\Delta\%$ (ε_p) = epsilon-constraint MIP per percentile; ReCANet $\Delta\%$ (λ) = ReCANet exponential weighting per λ.

	Recall					NDCG					Biz Value				
K	Orig	Mult $\Delta\%$	WS $\Delta\%$ (α)	ε $\Delta\%$ (ε_p)	ReCANet $\Delta\%$ (λ)	Orig	Mult $\Delta\%$	WS $\Delta\%$ (α)	ε $\Delta\%$ (ε_p)	ReCANet $\Delta\%$ (λ)	Orig	Mult $\Delta\%$	WS $\Delta\%$ (α)	ε $\Delta\%$ (ε_p)	ReCANet $\Delta\%$ (λ)
			0.1 −65.90%	90 −33.37%				0.1 −58.96%	90 −15.60%				0.1 +28.11%	90 +17.04%	
			0.2 −54.86%	80 −12.97%				0.2 −42.39%	80 −3.80%				0.2 +26.26%	80 +10.57%	
			0.3 −45.23%	70 −3.05%	1 +0.67%			0.3 −29.96%	70 −1.03%	1 +0.78%			0.3 +23.65%	70 +6.24%	1 +0.75%
			0.4 −29.42%	60 −1.60%	2 +0.69%			0.4 −18.71%	60 −0.19%	2 +0.13%			0.4 +19.82%	60 +3.34%	2 +1.16%
			0.5 −17.54%	50 0.00%	3 −1.20%			0.5 −10.42%	50 +0.06%	3 −1.40%			0.5 +15.50%	50 +1.76%	3 +0.74%
			0.6 −11.28%	40 −0.64%	4 −0.42%			0.6 −4.09%	40 −0.05%	4 −1.02%			0.6 +11.50%	40 +0.84%	4 +1.20%
			0.7 −5.22%	30 −1.52%	5 +0.52%			0.7 −1.90%	30 −0.48%	5 +0.02%			0.7 +8.10%	30 +0.28%	5 +1.47%
			0.8 −3.54%	20 −0.64%				0.8 −0.24%	20 −0.27%				0.8 +5.16%	20 +0.10%	
5	0.1215	−9.32%	0.9 −1.89%	10 −0.04%		0.1980	−5.23%	0.9 −0.62%	10 −0.04%		182.30	+9.41%	0.9 +2.37%	10 +0.02%	
			0.1 −59.90%	90 −30.01%				0.1 −57.05%	90 −17.29%				0.1 +25.77%	90 +17.20%	
			0.2 −53.36%	80 −10.88%				0.2 −45.49%	80 −5.23%				0.2 +24.77%	80 +10.38%	
			0.3 −46.42%	70 −6.06%	1 +1.24%			0.3 −35.42%	70 −2.22%	1 +0.96%			0.3 +22.81%	70 +5.43%	1 +0.32%
			0.4 −30.97%	60 −1.57%	2 +1.16%			0.4 −23.73%	60 −0.51%	2 +0.49%			0.4 +19.30%	60 +2.20%	2 +0.56%
			0.5 −18.25%	50 −0.29%	3 +0.29%			0.5 −14.01%	50 −0.16%	3 −0.90%			0.5 +14.98%	50 +0.76%	3 +0.44%
			0.6 −10.78%	40 +0.46%	4 +0.64%			0.6 −7.37%	40 +0.12%	4 −0.70%			0.6 +11.19%	40 +0.22%	4 +0.75%
			0.7 −5.33%	30 −0.03%	5 +1.91%			0.7 −4.27%	30 −0.03%	5 +0.04%			0.7 +7.96%	30 +0.04%	5 +0.84%
			0.8 −1.20%	20 −0.03%				0.8 −1.05%	20 −0.02%				0.8 +4.97%	20 +0.01%	
10	0.1711	−6.93%	0.9 −0.14%	10 0.00%		0.1952	−6.77%	0.9 −0.70%	10 0.00%		357.85	+7.11%	0.9 +2.41%	10 0.00%	
			0.1 −53.90%	90 −27.21%				0.1 −54.96%	90 −18.13%				0.1 +22.52%	90 +17.55%	
			0.2 −48.08%	80 −8.07%				0.2 −44.88%	80 −3.91%				0.2 +21.94%	80 +11.12%	
			0.3 −37.83%	70 −1.93%	1 +0.63%			0.3 −34.66%	70 −1.12%	1 +0.86%			0.3 +20.54%	70 +5.69%	1 +0.26%
			0.4 −22.80%	60 −0.23%	2 −0.46%			0.4 −22.91%	60 +0.02%	2 +0.33%			0.4 +17.91%	60 +1.85%	2 +0.43%
			0.5 −11.93%	50 −0.35%	3 +0.34%			0.5 −13.42%	50 −0.10%	3 −0.69%			0.5 +14.70%	50 +0.37%	3 +0.40%
			0.6 −5.54%	40 0.00%	4 +0.72%			0.6 −6.73%	40 0.00%	4 −0.18%			0.6 +11.45%	40 +0.05%	4 +0.59%
			0.7 −2.44%	30 0.00%	5 +1.17%			0.7 −3.74%	30 0.00%	5 +0.44%			0.7 +8.24%	30 +0.01%	5 +0.67%
			0.8 −0.39%	20 0.00%				0.8 −1.13%	20 0.00%				0.8 +5.29%	20 0.00%	
20	0.2150	−3.07%	0.9 −0.11%	10 0.00%		0.1955	−6.12%	0.9 −0.80%	10 0.00%		696.88	+5.37%	0.9 +2.59%	10 0.00%	
			0.1 −38.79%	90 −23.59%				0.1 −48.20%	90 −18.98%				0.1 +15.86%	90 +8.13%	
			0.2 −32.01%	80 −12.13%				0.2 −39.00%	80 −7.26%				0.2 +15.53%	80 +11.86%	
			0.3 −20.63%	70 −2.49%	1 +0.30%			0.3 −28.98%	70 −1.28%	1 +0.64%			0.3 +14.76%	70 +6.92%	1 +0.01%
			0.4 −12.05%	60 −0.67%	2 +0.99%			0.4 −19.30%	60 −0.16%	2 +0.64%			0.4 +13.61%	60 +2.16%	2 +0.17%
			0.5 −7.36%	50 +0.21%	3 +1.64%			0.5 −11.98%	50 +0.09%	3 −0.13%			0.5 +12.13%	50 +0.18%	3 +0.06%
			0.6 −5.09%	40 +0.06%	4 +1.16%			0.6 −6.80%	40 +0.02%	4 −0.10%			0.6 +10.38%	40 0.00%	4 +0.24%
			0.7 −2.94%	30 0.00%	5 +1.18%			0.7 −4.00%	30 0.00%	5 +0.48%			0.7 +8.34%	30 0.00%	5 +0.22%
			0.8 −0.91%	20 0.00%				0.8 −1.59%	20 0.00%				0.8 +6.01%	20 0.00%	
50	0.2773	−0.04%	0.9 +0.03%	10 0.00%		0.2101	−4.70%	0.9 −0.88%	10 0.00%		1641.52	+3.76%	0.9 +3.26%	10 0.00%	

9.3 Full Performance Comparison Tables

9.4 Pareto Frontier Plots

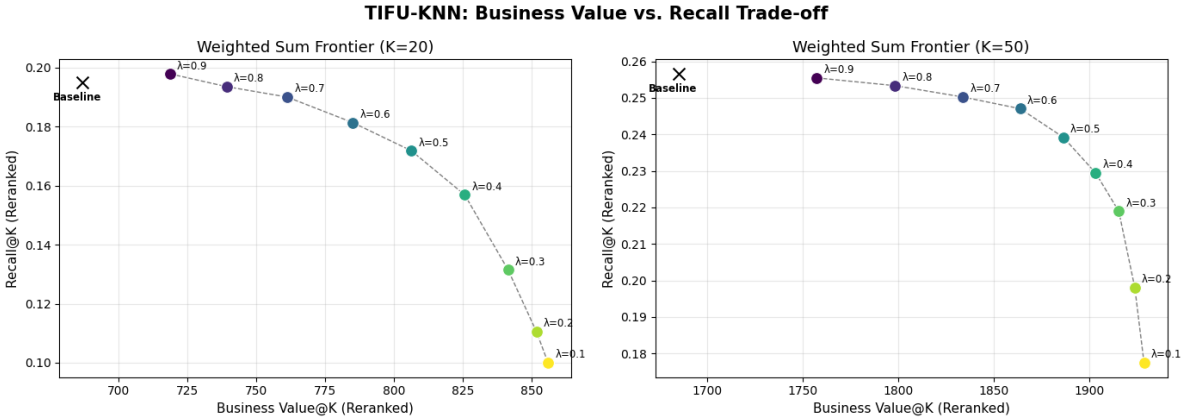

Figure 9.1: TIFU-KNN weighted sum Pareto frontier at $K = 20$ and $K = 50$.

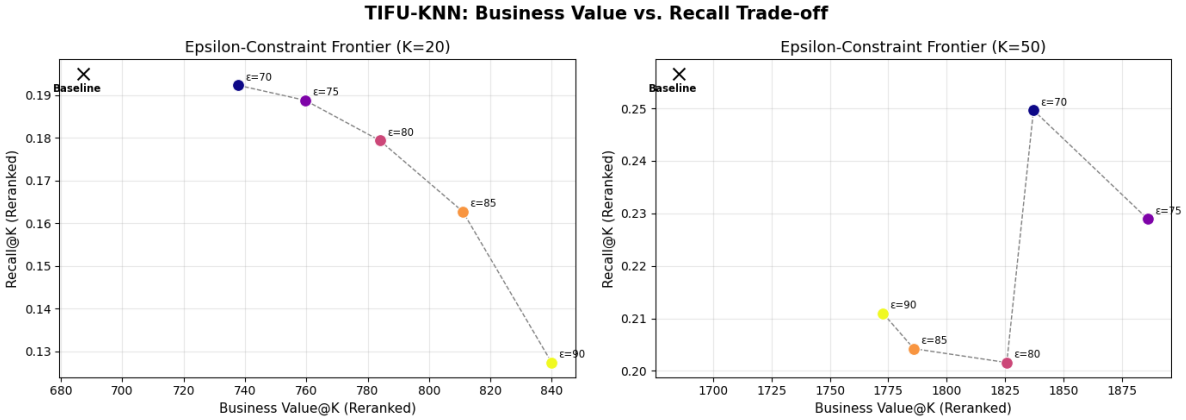

Figure 9.2: TIFU-KNN epsilon-constraint Pareto frontier at $K = 20$ and $K = 50$, $\varepsilon \in \{70, 75, 80, 85, 90\}$.

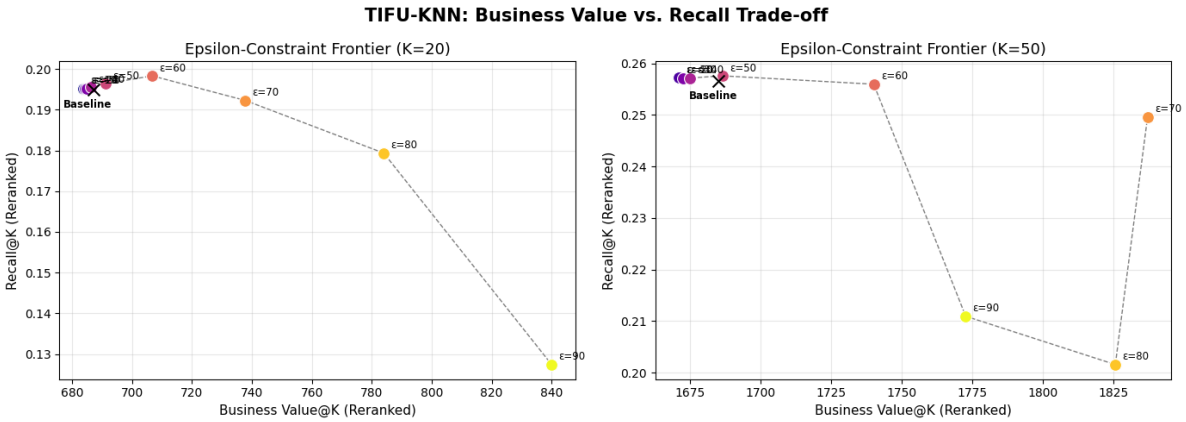

Figure 9.3: TIFU-KNN epsilon-constraint Pareto frontier at $K = 20$ and $K = 50$, full ε range.

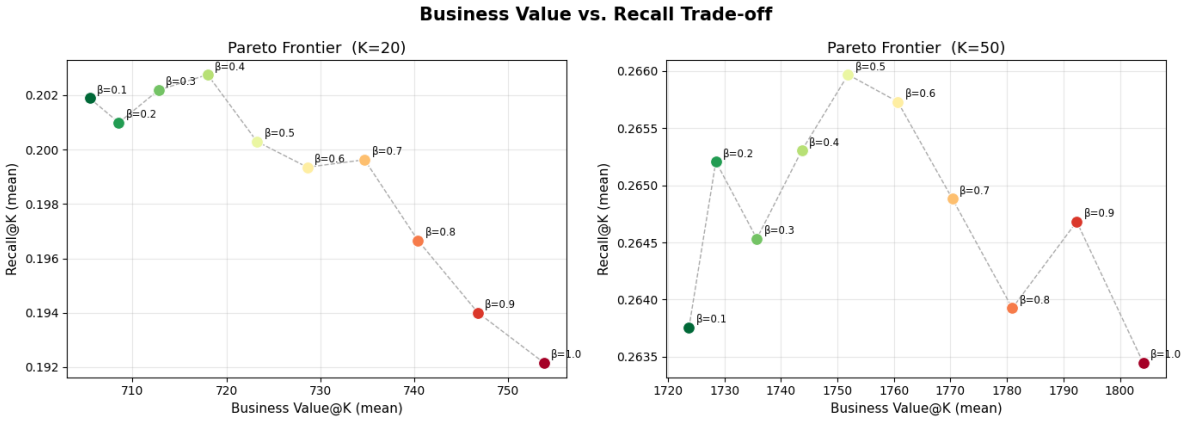

Figure 9.4: Adjusted TIFU-KNN Pareto frontier at $K = 20$ and $K = 50$, $\beta \in [0.1, 1.0]$.

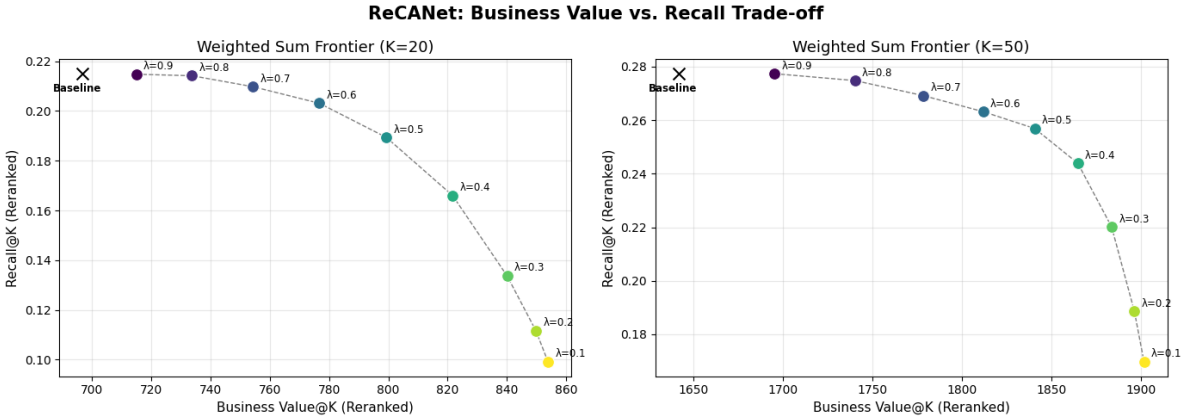

Figure 9.5: ReCANet weighted sum Pareto frontier at $K = 20$ and $K = 50$.

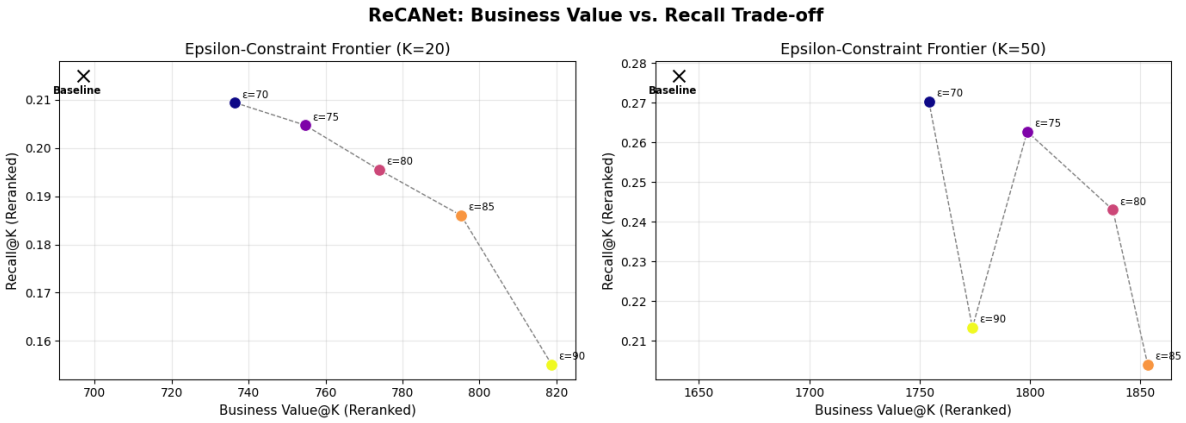

Figure 9.6: ReCANet epsilon-constraint Pareto frontier at $K = 20$ and $K = 50$, $\varepsilon \in \{70, 75, 80, 85, 90\}$.

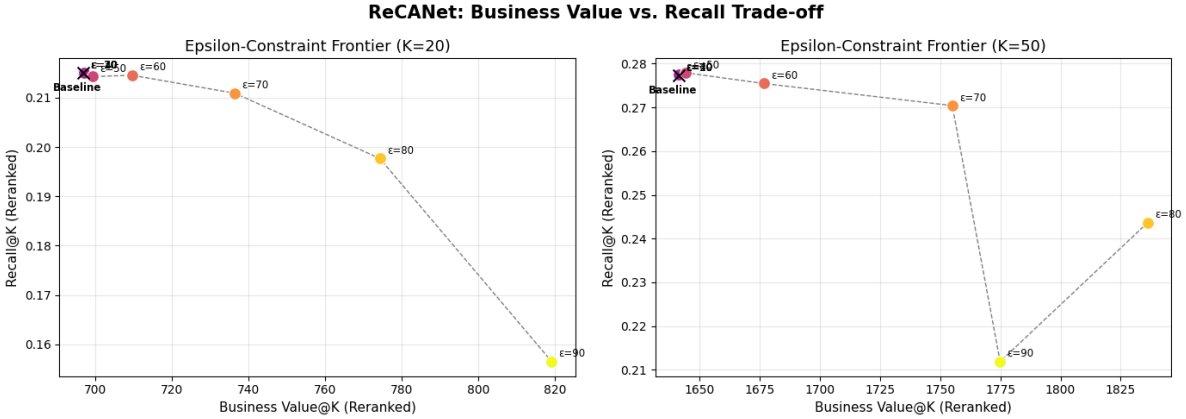

Figure 9.7: ReCANet epsilon-constraint Pareto frontier at $K = 20$ and $K = 50$, full ε range.