

HYBRID CLUSTERING AND RECOMMENDATION SYSTEM FOR E-COMMERCE CUSTOMER PERSONALIZATION

PROJECT PROPOSAL

Chien-Wei Weng & Ke Chen

Nicolas Perion-Quémeneur & Zihan Yang

CentraleSupélec, Université Paris-Saclay

MSc Data Sciences and Business Analytics

Email: chien-wei.weng@student-cs.fr

ke.chen@student-cs.fr

nicolas.perion-quemeneur@student-cs.fr

zihan.yang@student-cs.fr

1 INTRODUCTION

E-commerce recommendation systems typically treat all customers uniformly, despite considerable heterogeneity in shopping patterns and product preferences. Given the Instacart Online Grocery Shopping Dataset Instacart (2017) containing over 3 million orders from 200,000 users across 50,000 products, we use a two-phase approach combining customer segmentation (K-means, DBSCAN, Hierarchical clustering) with personalized recommendation systems (Collaborative Filtering, Content-Based, and Hybrid methods) to evaluate whether segment-aware recommendations outperform global one-size-fits-all approaches.

2 MOTIVATION AND PROBLEM DEFINITION

Current e-commerce recommendation systems predominantly employ global models that generate uniform recommendations for all users, regardless of shopping behavior heterogeneity. This one-size-fits-all approach limits recommendation relevance. For example, a frequent organic produce shopper and an occasional bulk buyer receive similar product suggestions, despite totally different purchasing patterns and needs.

We investigate whether customer segmentation can improve recommendation accuracy. Specifically, we first cluster users into behavioral groups using RFM metrics (Recency, Frequency, Monetary value), basket size, and temporal shopping patterns, then train specialized recommendation models for each segment. Our hypothesis is that segment-specific models will outperform global approaches by better capturing group-level preferences.

This problem matters for three reasons. First, personalization drives 10–15% of e-commerce revenue McKinsey (2021), yet most systems still use global models. Second, segment-based approaches offer interpretability: retailers can understand *which* customer groups exist and *why* certain recommendations work for specific segments. Third, practical applications include targeted marketing campaigns, inventory optimization aligned with segment preferences, and dynamic pricing strategies leveraging segment-specific price sensitivities.

Recent research shows promise for segment-based approaches. Emre et al. (2023) demonstrated improved precision using customer segmentation in fashion retail, while systematic reviews Miguel et al. (2023); Ilham et al. (2024) identify K-means clustering and hybrid recommendation methods as most effective for e-commerce. However, no prior work systematically compares multiple clustering algorithms and improved customer retention strategies based on segment-specific behavior patterns.

We fill this gap by evaluating three clustering algorithms (K-means, DBSCAN, Hierarchical) and four recommendation strategies (Baseline, Collaborative Filtering, Content-Based, Hybrid) to quantify segment-specific versus global model performance.

3 METHODOLOGY

3.1 OVERVIEW AND DATA PROCESSING

Our methodology consists of two phases: customer segmentation followed by personalized recommendation. We use the Instacart Online Grocery Shopping Dataset containing over 3 million orders from 200,000 users across 50,000 products. Using the complete Instacart dataset, we apply temporal splitting where each user’s most recent order becomes the test set and all prior orders form the training set. This temporal approach is more appropriate than random splitting for recommendation systems, as it preserves the sequential nature of purchase behavior. All features for segmentation are calculated exclusively on training data to prevent information leakage. We normalize features using StandardScaler before clustering, which is essential for distance-based algorithms.

3.2 PHASE 1: CUSTOMER SEGMENTATION

3.2.1 FEATURE ENGINEERING

This project constructs user-level features capturing three dimensions of shopping behavior. First, we calculate RFM metrics Anitha et al. (2022): Recency (days since last order), Frequency (total number of orders), and Monetary value (total items purchased, serving as a spending proxy since actual prices are unavailable). Second, we extract behavioral features including average basket size, reorder ratio (percentage of previously purchased items), and percentage distribution across product departments and aisles, which captures category preferences. Third, we incorporate temporal patterns: preferred shopping day of week, preferred hour of day, and average interval between orders.

3.2.2 CLUSTERING ALGORITHMS

We compare three clustering approaches representing different algorithmic paradigms Miguel et al. (2023). K-means clustering uses partition-based optimization to minimize within-cluster variance, making it efficient for large datasets Anitha et al. (2022). DBSCAN employs density-based clustering to identify arbitrarily shaped clusters while handling noise and outliers Adi et al. (2024). Hierarchical clustering builds a dendrogram through agglomerative merging, providing interpretable cluster structure Miguel et al. (2023).

3.2.3 EVALUATION AND SELECTION

We assess clustering quality and determine the optimal number of clusters using three standard practices: elbow method, Silhouette Score, which measures cluster cohesion and separation (higher is better), and Davies-Bouldin Index, which evaluates cluster separation quality by measuring the ratio of within-cluster to between-cluster distance (lower is better) Miguel et al. (2023); Adi et al. (2024). We select the optimal clustering algorithm and number of segments based on these metrics combined with qualitative assessment of segment interpretability.

3.3 PHASE 2: RECOMMENDATION SYSTEMS

3.3.1 RECOMMENDATION APPROACHES

We implement and compare four recommendation strategies. The Baseline approach uses popularity-based recommendation, ranking products by global purchase frequency. Collaborative filtering Ilham et al. (2024) employs singular value decomposition (SVD) to learn latent factors from the user-item interaction matrix, predicting purchase likelihood based on similar user behaviors. Content-Based filtering Ilham et al. (2024) computes item similarity using product features (department and aisle categories), recommending items similar to each user’s purchase history through cosine similarity of one-hot encoded category vectors. The Hybrid approach combines collaborative and content-based methods through weighted averaging: $\text{score}_{\text{hybrid}} = \alpha \times \text{score}_{\text{CF}} + (1 - \alpha) \times \text{score}_{\text{content}}$, where we experiment with $\alpha \in \{0.5, 0.7, 0.9\}$ and report results for the best-performing weight Widayanti et al. (2023).

3.3.2 EVALUATION STRATEGY

We follow a two-stage evaluation process. First, we train all four approaches as global models using data from all users. We evaluate each model using Precision@K, Recall@K, and F1@K for $K \in \{5, 10, 20\}$, standard metrics for recommendation systems Ilham et al. (2024); Emre et al. (2023). We select the best-performing approach based on these metrics. Second, we train the winning approach separately for each customer segment, creating segment-specific models. For users in the test set, we assign them to their corresponding segment and generate recommendations using that segment’s specialized model. We then compare segment-specific model performance against the global model to quantify the benefit of segmentation-based personalization.

3.4 IMPLEMENTATION DETAILS

We implement our approach in Python using standard libraries: scikit-learn for clustering, evaluation metrics, and cosine similarity computation; scipy for hierarchical clustering; and Surprise for collaborative filtering. We use SVD with default configuration (100 latent factors) to avoid hyperparameter tuning overhead, as our focus is on comparing recommendation approaches rather than optimizing individual models. For content-based filtering, we one-hot encode product categories (21 departments and 134 aisles, yielding 155 binary features per product) and compute pairwise cosine similarities using scikit-learn. Parameters are selected based on preliminary experiments. Note that k-fold cross-validation is not employed, as temporal ordering is critical for realistic evaluation in recommendation systems.

4 EVALUATION

4.1 EXPERIMENTAL SETUP

This project evaluates the proposed approach using the complete Instacart dataset with temporal splitting, where approximately 200,000 users contribute test data (one order per user). The training set comprises users’ historical purchase sequences, while the test set contains their most recent orders for validation.

4.2 EVALUATION METRICS

For clustering evaluation, we use elbow method, Silhouette Score, and Davies-Bouldin Index as described in Section 3. For recommendation quality, we employ three standard metrics Ilham et al. (2024); Emre et al. (2023). Precision@K measures the proportion of recommended items that the user actually purchased. Recall@K measures the proportion of purchased items that were successfully recommended. F1@K provides the harmonic mean of precision and recall, balancing both metrics. We evaluate at $K \in \{5, 10, 20\}$ to represent different recommendation scenarios.

4.3 EVALUATION PROTOCOL

First, we select the optimal clustering algorithm using clustering metrics. Second, we train and compare all four recommendation approaches (including baseline) as global models, then train segment-specific versions of the best-performing approach and compare against both the global version and baseline using paired t-tests on per-user recommendation performance.

4.4 SUCCESS CRITERIA

We consider segmentation-based personalization successful if segment-specific models outperform the global model on the majority of evaluation metrics (at least 5 out of 9 metric-K combinations). Results will be presented in tables comparing global versus segment-specific performance across all metrics and K values.

REFERENCES

- S. P. Adi et al. Comparison of k-means and dbscan algorithms for customer segmentation in e-commerce. *Journal of Digital Market and Digital Currency*, 2024. URL <https://dspace.uc.ac.id/handle/123456789/8154>. Citations: 21.
- P. Anitha et al. Rfm model for customer purchase behavior using k-means algorithm. *Journal of King Saud University – Computer and Information Sciences*, 2022. URL <https://www.sciencedirect.com/science/article/pii/S1319157819309802>. Citations: 337.
- Y. Emre et al. A hyper-personalized product recommendation system focused on customer segmentation: An application in the fashion retail industry. *Journal of Theoretical and Applied Electronic Commerce Research*, 2023. URL <https://www.mdpi.com/0718-1876/18/1/29>. Citations: 52.
- S. Ilham et al. Systematic literature review on recommender system: Approach, problem, evaluation techniques, datasets. *IEEE*, 2024. URL <https://ieeexplore.ieee.org/abstract/document/10415424>. Citations: 52.
- Instacart. The instacart online grocery shopping dataset 2017, 2017. URL <https://www.kaggle.com/datasets/psparks/instacart-market-basket-analysis/data>.
- McKinsey. The value of getting personalization right—or wrong—is multiplying. *McKinsey Digital*, 2021. URL <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying>.
- A. G. Miguel et al. A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and e-Business Management*, 2023. URL <https://link.springer.com/article/10.1007/s10257-023-00640-4#Sec13>. Citations: 157.
- R. Widayanti et al. Improving recommender systems using hybrid techniques of collaborative filtering and content-based filtering. *Journal of Applied Data Sciences*, 2023. URL <https://bright-journal.org/Journal/index.php/JADS/article/view/115>. Citations: 137.