Art of Suggestion: Building Personalized Recommendation Engines for Instacart

Akansha Lalwani
Halıcıoğlu Data Science Institute
University of California, San Diego
alalwani@ucsd.edu

Saksham Jindal
Department of Electrical and Computer Engineering
University of California, San Diego
sjindal@ucsd.edu

***Abstract*—This report presents a comprehensive analysis of customer behavior and the development of a recommender system for Instacart, a leading online grocery delivery platform. Utilizing a rich dataset from Instacart, this study employs collaborative filtering techniques, including TF-IDF Neighborhood, Singular Value Decomposition (SVD), and Bayesian Personalized Ranking (BPR), to predict user preferences and improve product recommendations. Our evaluation focuses on the Mean Recall metric to assess the effectiveness of these models. The results indicate a significant improvement in recommendation quality compared to a baseline model, highlighting the potential of personalized recommendation systems in enhancing the online shopping experience. This research contributes to the growing field of e-commerce recommender systems, offering insights into customer purchasing patterns and the efficacy of different recommendation algorithms.**

I. Introduction

The rise of online platforms for food and grocery shopping has transformed the way people access their favorite products. This shift, further accelerated by the 2020 coronavirus pandemic, has added to exponential growth of online delivery services.

Established in 2012, Instacart is an online service providing delivery and curbside pickup across 5,500+ cities in the United States and Canada[3]. Customers can easily place grocery orders through the website or mobile app, with an 'Instacart Shopper' handling the pickup and delivery. Instacart has partnerships with major local and national market chains, including Vons, Walmart, and BestBuy.

This report aims to understand customer spending habits on groceries and provide recommendations based on these patterns. Recognizing these habits is crucial for businesses to effectively market products and influence consumer choices. With continued growth expected in grocery and supermarket industry, an efficient recommendation system can give companies a competitive advantage and boost profits.

***Index Terms*—e-commerce, recommender systems, collaborative filtering**

II. About the Dataset

The Instacart dataset consists of six dataset files that describe customer order history. These files, named *orders*, *products*, *order_products_prior*, *order_products_train*, *aisles*, and *departments*, collectively form a relational set of information.

A. Orders

The `orders` dataset contains 3.4 million entries with seven features, including `user_id`, order history, order placement details, and days since the last order. Each observation corresponds to an Instacart grocery order.

B. Reorders

The `reorders` dataset, combining prior and training reorders, includes 33.8 million entries with information on product addition order, `product_id`, `add_to_cart_order`, and a binary indicator for product reordering.

C. Products

Containing 49,688 entries, the `products` dataset outlines unique product details, including `product_id`, `product_name`, `aisle_id`, and `department_id`.

D. Aisles and Departments

The `aisles` dataset (134 entries) names different aisles, while the `departments` dataset (21 entries) provides department names. Both datasets offer identification numbers for aisles and departments, along with their respective names. No single aisle overlaps with multiple departments.

This dataset structure enables a comprehensive exploration of customer behaviors and product interactions for the subsequent analyses in this study.

III. Dataset Exploratory Analysis

Our exploratory analysis aims to uncover underlying patterns and relationships within the Instacart dataset. This section provides an overview of our initial findings and observations which informed our predictive modeling approach.

A. Orders by Aisles

In Figure 1, we observe that fruits and vegetable aisles are the most popular aisles among the shoppers. This is followed by aisles containing dairy products are the most popular.

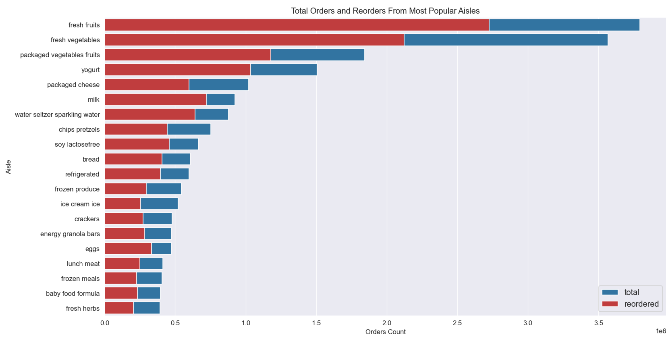

Fig. 1: Total Orders vs Aisles

B. Orders by Departments

We observe that produce and dairy egss are the most popular product categories and form 29 % and 17 % of all the orders (Figure 2). This is in line with the fruits and vegetables being the most popular aisles.

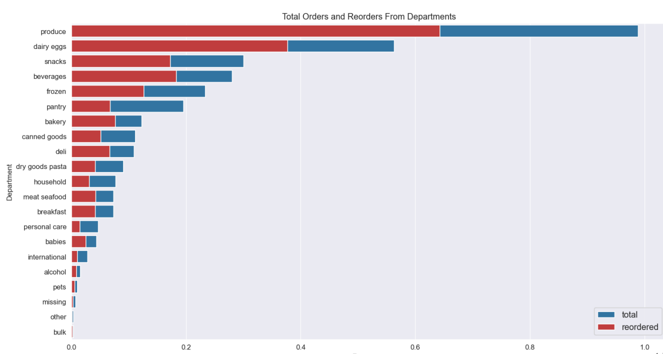

Fig. 2: Total Orders vs Departments

C. Orders by Products

It is interesting to observe that bananas (inorganic and organic) are the most popular products (Figure 3). We also observe that organic fruits and vegetables are more popular than their inorganic counterparts. We also plot unique users (Figure 4) of most popular products.

D. Orders by time

We observe that the shoppers place most orders on Sundays and least orders on Thursday (Figure 6). We speculate that most shoppers tend to plan their groceries before the week starts and use their groceries during the week.

IV. Predictive Task and Evaluation

A. Definition of the Predictive Task

The primary goal of our predictive task is to develop a recommender system capable of effectively suggesting products to Instacart users based on their historical purchase data.

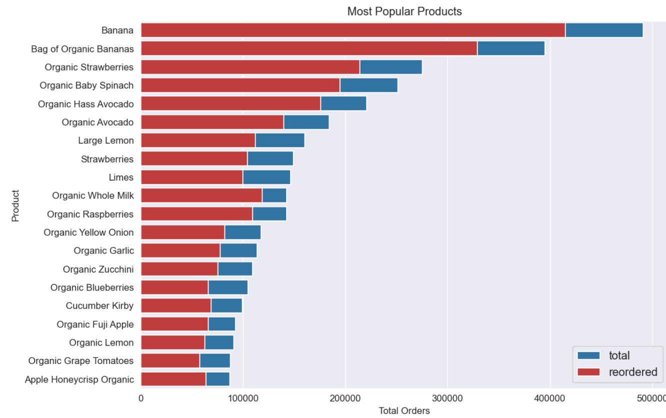

Fig. 3: Total orders of most popular products

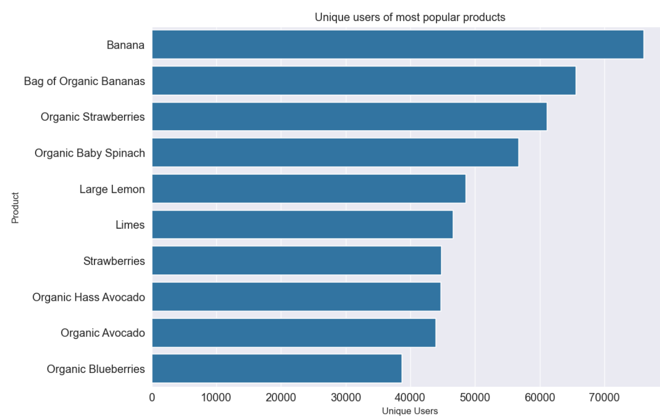

Fig. 4: Unique users of most popular products

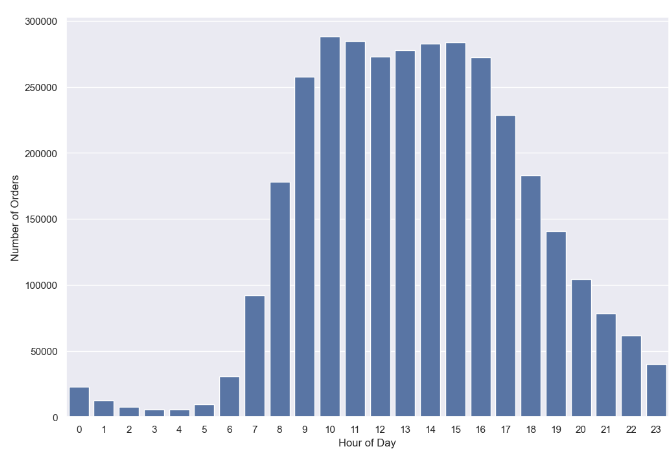

Fig. 5: Number of Orders vs Hours of the day

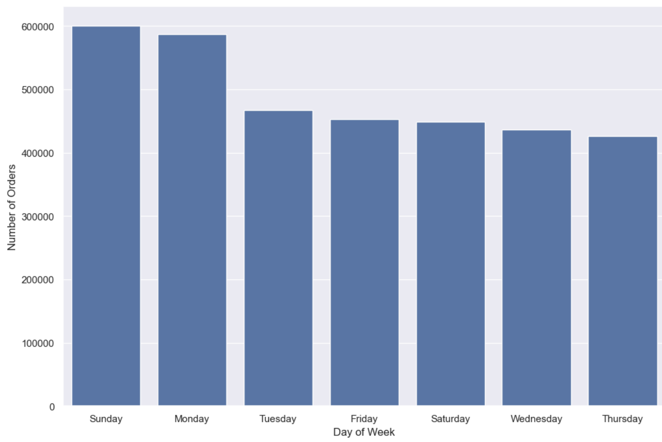

Fig. 6: Number of Orders vs Days of the week

Unlike basic recommendation models that might suggest popular items or random selections, our system aims to identify patterns in user behavior, leveraging these insights to predict future purchases with a high degree of personalization.

Specifically, the task involves:

1) Analyzing user purchase histories to identify frequent item combinations and preferred products.
2) Developing algorithms that can predict the likelihood of a product being purchased by a user in their next order.
3) Comparing the performance of our recommender system against a baseline model to validate the effectiveness of our approach.

B. Predictive Task Formulation

Let $\mathcal{U}$ be the set of users, $\mathcal{P}$ the set of products, and $\mathcal{R}$ the user-product interaction matrix, where $R_{u,p}$ represents the interaction of user u with product p. Our goal is to predict the missing entries in $\mathcal{R}$, i.e., $R_{u,p}$ for user-product pairs (u,p) where $R_{u,p}$ is unknown.

Mathematically, the predictive task can be expressed as follows:

$$\hat{R}_{u,p} = f(u,p)$$

where $\hat{R}_{u,p}$ is the predicted interaction score for user u and product p, and $f(u,p)$ is the prediction function.

C. Data Preparation and Feature Utilization

We undertook the following steps in preparing our dataset:

1) **Data Collection**: Utilization of the 'Instacart Online Grocery Shopping Dataset 2017', featuring extensive order histories from a wide user base.
2) **Data Integrity**: Verification of the dataset's completeness and accuracy, ensuring each user's history was substantial and consistent.
3) **Data Structuring**: Segmentation of the dataset into historical and recent purchases, forming the backbone of our user-product interaction matrix.

D. Evaluation Framework

For an effective evaluation of our recommender system models, we adopted the following framework:

1) **Test Dataset**: Utilization of a separate dataset comprising current purchases of approximately 26,000 users as our test set.
2) **Baseline Model**: The effectiveness of our models was compared against a baseline model that randomly recommends products to users, serving as a simplistic benchmark.
3) **Assessment Metrics**: Success of our models was measured by their capability to predict actual user purchases compared to the random baseline.

Through this framework, we sought not only to demonstrate the superiority of our models over basic random recommendations but also to provide insights into the nuances of user behavior and preferences.

E. Business Implications

The implementation of our recommender system is projected to have a significant impact on Instacart's business:

1) **Sales Conversions**: By aligning product suggestions with individual user preferences, the likelihood of purchase conversions is increased, thereby boosting sales.
2) **Market Insight**: The predictive task will generate valuable insights into market trends and consumer behaviors, informing strategic business decisions.
3) **Customer Retention**: Personalized experiences foster customer loyalty, encouraging repeated use of the Instacart platform.

F. Enhancing User Experience

The recommender system is designed to significantly improve the Instacart shopping experience:

1) **Personalization**: Users will receive recommendations tailored to their unique tastes and preferences, creating a more intuitive shopping experience.
2) **Convenience**: The system's ability to anticipate user needs and suggest relevant products will streamline the shopping process, saving users time and effort.
3) **Discovery**: By introducing users to products they may like but have not yet encountered, the system will enhance the joy of discovery during the shopping journey.

V. Model Selection and Approach

Our study employed three collaborative filtering methods, each chosen for its specific suitability to the Instacart dataset and our predictive goals.

A. Model Description

1) **TF-IDF Based Neighborhood Approach**
This model capitalizes on the similarities in purchase behaviors among users, offering a tailored recommendation experience.

Mathematically, the predicted interaction score can be computed as:

$$\hat{R}_{u,p} = \sum_{v \in N(u)} \text{TF-IDF}(u,v) \cdot R_{v,p}$$

where $N(u)$ is the neighborhood of users similar to u based on TF-IDF scores.

2) **Singular Value Decomposition (SVD)**
SVD was applied to overcome the challenge of sparse data in our utility matrix, thereby uncovering hidden user-product relationships.
Mathematically, the predicted interaction score can be expressed as:

$$\hat{R}_{u,p} = \sum_{k=1}^{K} U_{u,k} \cdot \Sigma_{k,k} \cdot V_{p,k}$$

where U, Σ, and V are the matrices obtained from the SVD decomposition, and K is the chosen number of latent factors.

3) **Bayesian Personalized Ranking (BPR)** Excelling in datasets with implicit feedback, BPR enhances the personal ranking of items by comparing pairs of products based on user interactions. This method proved especially suitable for our Instacart dataset, enabling inference of user preferences without explicit ratings.
The pairwise ranking objective of BPR can be defined as:

$$\mathcal{L} = -\sum_{(u,i,j) \in \mathcal{D}} \log \sigma(\hat{R}_{u,i} - \hat{R}_{u,j})$$

where $\mathcal{D}$ is the set of user-item pairs, and σ is the sigmoid function.

B. Comparison of Model Strengths and Weaknesses

The TF-IDF based Neighborhood Model emerges as a robust approach, providing transparent and interpretable personalized recommendations by leveraging user similarities in purchase behaviors. However, its effectiveness is contingent upon accurate TF-IDF calculations, making it susceptible to noise and outliers. Additionally, the model faces challenges in addressing the cold-start problem, particularly for new users or products with limited purchase history.

Singular Value Decomposition (SVD) stands out for its ability to handle sparse data effectively, capturing latent factors that enable accurate predictions even for users with limited interaction history. Despite this, scalability issues may arise with larger datasets, necessitating dimensionality reduction techniques. Moreover, the abstract nature of latent factors compromises the interpretability of the model.

The Bayesian Personalized Ranking (BPR) model proves advantageous in scenarios with implicit feedback, offering a solution to the cold-start problem by focusing on pairwise comparisons. However, its trade-off involves a sacrifice in interpretability, as the model's workings become less transparent. Additionally, BPR demands careful hyperparameter tuning for optimal performance.

In summary, the choice among these collaborative filtering models depends on the specific characteristics of the dataset and the desired balance between interpretability and predictive performance. The TF-IDF based Neighborhood Model offers transparency, SVD excels in handling sparsity, and BPR addresses implicit feedback scenarios, each with its own set of strengths and limitations. In our case, while BPR is potent for implicit datasets like Instacart's, it may struggle with cold-start scenarios where user data is limited.

C. Optimization and Challenges

In optimizing the models, we aimed to strike a balance between accuracy and computational efficiency. We addressed scalability concerns through efficient data structuring techniques and fine-tuned model parameters to prevent overfitting. The consideration of alternative models, including deep learning-based recommenders, was part of our exploration. However, these were ultimately excluded due to their complexity and resource-intensive nature, which could have hindered practical implementation and limited the feasibility of our study. This decision aligns with our focus on achieving a practical and efficient solution for personalized recommendations in the context of the Instacart dataset.

VI. Related Literature

Our project on developing a recommender system for Instacart intersects with a vibrant area of research in e-commerce. This section delves into relevant literature, including dataset origins, similar studies, and cutting-edge methods in recommender systems.

A. Instacart Dataset from Kaggle

The Instacart dataset, pivotal to our analysis, is accessible on Kaggle [1]. It encompasses over 3 million orders from over 200,000 users, offering a comprehensive view of online grocery shopping behavior. Previous research using this dataset has focused on aspects like purchase prediction, product bundling, and user segmentation, providing valuable insights into consumer patterns.

B. Similar Datasets and Studies

Databases like the MovieLens dataset [2] and the Amazon product dataset [3] have undergone extensive scrutiny. Studies using these datasets have explored personalized recommendation systems, often using complex algorithms to understand user preferences and predict product ratings.

C. State-of-the-Art Methods in Recommender Systems

The field has witnessed significant advancements, notably:

1) **Deep Learning Techniques**:
   - "Neural Collaborative Filtering" by Xiangnan He et al. [3] introduces a deep learning approach to collaborative filtering, achieving state-of-the-art performance.

- ”Variational Autoencoders for Collaborative Filtering” by Dawen Liang et al. [4] proposes a probabilistic approach using variational autoencoders to model user preferences effectively.
- ”Graph Neural Networks for Collaborative Filtering” by Thomas N. Kipf and Max Welling [5] suggests a graph neural network approach, particularly useful for modeling user-item interactions with a graph structure.

2) **Hybrid Recommender Systems**:
   - ”A Hybrid Recommender System Using Collaborative Filtering and Content-Based Filtering” by Robin Burke [6] combines collaborative and content-based filtering for more nuanced recommendations.
   - ”A Hybrid Recommender System with Pluralistic Meta-Learning” by Hongzhi Yin et al. [7] leverages meta-learning to enhance performance by combining multiple recommendation algorithms.
   - ”A Hybrid Recommender System Combining Multi-View Collaborative Filtering and Content-Based Filtering” by Xiaolin Qu et al. [8] integrates multi-view collaborative filtering and content-based filtering for a richer recommendation strategy.

3) **Context-Aware Recommendations**:
   - ”PowerMat” by Tefagh et al. [9] addresses the cold-start problem in context-aware systems, a challenge relevant to our study.

D. Comparison with Existing Work

Our findings resonate with and build upon existing research, though our focus on the Instacart dataset introduces unique insights into grocery e-commerce consumer behavior. While aligning with recent trends, our study also addresses gaps such as the cold-start problem, contributing to a more nuanced understanding of e-commerce recommender systems.

In summary, our literature review situates our project within the ongoing discourse in e-commerce recommender systems, offering a benchmark for comparing our methods and findings with the state-of-the-art.

VII. Results and Evaluation

A. Evaluation Methodology

Our evaluation of the Instacart recommender system centered around a recall-based metric, considering the dataset’s nature where a non-purchase doesn’t necessarily imply a dislike. We employed the mean recall, a metric that quantifies the proportion of recommended products that match actual purchases, as follows:

$$\text{Mean Recall} = \frac{|\{\text{recommended}\} \cap \{\text{actual}\}|}{|\{\text{actual}\}|}$$

This measure effectively gauges the alignment of our recommendations with user buying behavior.

B. Comparative Analysis of Methods

We observed the following results when comparing the performance of different recommendation techniques against a random baseline model:

Method	Mean Recall (%)	Run Time (mins)
Random Baseline	1.3	-
TF-IDF Neighborhood	19.5	166
MF using SVD (50 factors)	3.1	15
MF using SVD (100 factors)	3.0	17
BPR from Implicit Feedback	5.2	2

TABLE I: Comparative Results of Recommender System Methods

C. Performance Insights

The TF-IDF Neighborhood model demonstrated a substantial lead over the random baseline, indicating its strength in capturing user preferences based on purchasing patterns. Its ability to factor in the uniqueness and frequency of product purchases makes it a robust tool for personalized recommendations.

The BPR model, tailored for datasets with implicit feedback like Instacart’s, showed a notable improvement over the baseline. Its focus on pairwise ranking optimization, deducing preferences from user-item interactions, underscores its effectiveness in scenarios where explicit user feedback is absent.

Overall, the results reinforce the superiority of tailored recommendation models over basic random suggestions, emphasizing the need for nuanced approaches in understanding and predicting consumer behavior.

VIII. Conclusion and Future Directions

Our analysis of Instacart’s customer data demonstrates the significant impact of collaborative filtering methods, especially the TF-IDF Neighborhood and Bayesian Personalized Ranking (BPR), in recommender systems. The TF-IDF model excelled due to its nuanced approach to user purchase history, outperforming the baseline random recommendation model. Its ability to prioritize unique and rare purchases allows for personalized recommendations, underscoring its effectiveness in capturing consumer preferences.

BPR also showed promising results, particularly in contexts with implicit feedback, common in e-commerce datasets. Its performance, while not surpassing TF-IDF, was notably superior to the baseline, affirming its relevance in scenarios lacking explicit user feedback.

The success of these models can be attributed to their sophisticated handling of user-product interactions. The TF-IDF model’s parameters, emphasizing purchase frequency and uniqueness, provided insights into user preferences more accurately than simpler models. BPR’s effectiveness lies in its capacity to infer preferences from implicit user behaviors.

Looking ahead, addressing the cold-start problem remains a significant challenge. Future research should focus on hybrid models that blend collaborative and content-based filtering, possibly incorporating deep learning for enhanced personalization and accuracy. This approach could leverage richer user and product data, offering more refined recommendations in the evolving landscape of online grocery shopping.

In summary, our findings highlight the potential of advanced recommender systems in improving user experience and business strategies in e-commerce. Continuing to develop these systems will be crucial for adapting to the dynamic preferences of online consumers.

REFERENCES

[1] "The Instacart Online Grocery Shopping Dataset 2017", Accessed from Kaggle, [Online]. Available: https://www.kaggle.com/c/instacart-market-basket-analysis

[2] F. Maxwell Harper and Joseph A. Konstan, "The MovieLens Datasets: History and Context", ACM Transactions on Interactive Intelligent Systems (TiiS) 5, no. 4: 19:1–19:19, 2016, doi:10.1145/2827872.

[3] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua, "Neural Collaborative Filtering", in Proceedings of the 26th International Conference on World Wide Web, 2017, pp. 173–182, doi:10.1145/3038912.3052569.

[4] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara, "Variational Autoencoders for Collaborative Filtering", in Proceedings of the 2018 World Wide Web Conference on World Wide Web, 2018, pp. 689–698, doi:10.1145/3178876.3186150.

[5] Thomas N. Kipf and Max Welling, "Graph Neural Networks for Collaborative Filtering", arXiv:1609.02907 [cs.IR], September 2016, [Online]. Available: https://arxiv.org/abs/1609.02907.

[6] Robin Burke, "Hybrid Recommender Systems: Survey and Experiments", User Modeling and User-Adapted Interaction 12, 4 (2002): 331–370, doi:10.1023/A:1021240730564.

[7] Hongzhi Yin, Bin Cui, Hao Wang, Zhiyong Cheng, Xing Xie, and Xiaofang Zhou, "A Hybrid Recommender System with Pluralistic Meta-Learning", in Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1409–1418, doi:10.1145/3357384.3357938.

[8] Xiaolin Qu, Jian Zhou, and Zhiguo Gong, "A Hybrid Recommender System Combining Multi-View Collaborative Filtering and Content-Based Filtering", in Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019, pp. 3786–3792, doi:10.24963/ijcai.2019/525.

[9] Mehdi Tefagh, Hamed Jelodar, Hamid Beigy, and Ehsan Adeli, "PowerMat: Context-aware Recommender System without User Item Rating Values that Solves the Cold-start Problem", arXiv:2303.06356 [cs.IR], March 2023, DOI:10.48550/arXiv.2303.06356, [Online]. Available: https://arxiv.org/abs/2303.06356.