PERSONALIZED RECOMMENDATION SYSTEMS FOR ONLINE RETAIL USING SIMPLE MACHINE LEARNING TECHNIQUES

[1]Haritha.K.B, [2]G.K.Karthika

Assistant Professor, Research Scholar - Department of Commerce, Sri Ramakrishna College of Arts and Science for Women, Coimbatore- 641044

Assistant Professor - Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore - 641006

Abstract

The rapid expansion of online retail platforms has resulted in information overload, where customers are exposed to a very large number of products, making it difficult to identify items aligned with individual preferences. Personalized recommendation systems address this challenge by analyzing user behavior and historical interactions to generate relevant product suggestions. This study proposes a simple and computationally efficient recommendation framework based on user-based collaborative filtering implemented using the k-Nearest Neighbors algorithm. The proposed approach is designed for small and medium-sized online retailers with limited technical resources and focuses on interpretability, low computational cost, and ease of deployment. The experimental analysis conducted on an e-commerce interaction dataset demonstrates that the proposed method significantly outperforms a popularity-based recommendation baseline in terms of Precision@K and Recall@K. The findings suggest that even simple machine learning techniques can enhance customer engagement and improve personalization quality in online retail environments, providing a practical and scalable solution for commerce-driven digital platforms.

Keywords: Recommendation systems; E-commerce; Machine learning; Collaborative filtering; k-Nearest Neighbors; Personalization

1. Introduction

Online retail platforms have transformed modern commerce by providing customers with convenient access to extensive product catalogs; however, this abundance of options often leads to decision fatigue and reduced user satisfaction due to the difficulty of discovering relevant products. Large-scale platforms such as Amazon and Flipkart have demonstrated that personalized recommendation systems are central to improving user experience, increasing conversion rates, and enhancing customer retention through tailored product suggestions derived from historical user interactions. While advanced deep learning approaches have gained popularity in industrial-scale recommender systems, many small and medium online retailers face constraints related to limited datasets, restricted computational resources, and the lack of specialized technical expertise, which restricts the adoption of complex models. Consequently, there is a strong practical need for simple, interpretable, and cost-effective recommendation techniques that can deliver measurable commercial value without requiring extensive infrastructure. This study addresses this need by proposing and evaluating a basic machine learning–based recommendation framework that leverages collaborative filtering with k-Nearest Neighbors to provide personalized product recommendations in online retail settings.

2. Objectives

Early research on recommendation systems focused on memory-based collaborative filtering methods that compute similarity between users or items based on historical ratings or interactions, offering simplicity and interpretability. Content-based recommendation approaches rely on matching user preferences with product attributes and textual descriptions, while hybrid approaches combine collaborative and content-based methods to address limitations such as data sparsity and the cold-start problem. Although matrix factorization and deep learning methods have shown strong performance on large-scale datasets, their computational cost, complexity, and limited transparency pose challenges for smaller organizations. Recent studies emphasize that well-tuned simple models can achieve competitive performance in practical e-commerce environments, particularly when data quality is high and evaluation metrics are aligned with business objectives. This work builds upon these findings by focusing on a lightweight and interpretable approach that can be easily implemented in resource-constrained online retail systems.

3. Problem Statement

Small and medium online retailers often lack the infrastructure required to deploy complex recommendation systems, leading to reliance on non-personalized strategies such as popularity-based or manually curated recommendations, which fail to capture individual user preferences. The main challenges include limited computational resources, sparse user–item interaction data, the cold-start problem for new users and products, and the need for transparent recommendation logic that can be easily understood and maintained by developers and business stakeholders. The objective of this study is to design and evaluate a simple machine learning–based recommendation system that improves personalization quality while remaining computationally efficient, interpretable, and suitable for real-world deployment in small-scale online retail platforms.

2. Literature Review (with in-text citations)

The development of recommendation systems has received extensive research attention due to their strong commercial relevance in online retail environments, where personalization directly influences user engagement, satisfaction, and sales performance [1,2]. Early studies primarily focused on memory-based collaborative filtering methods, which generate recommendations by computing similarity between users or items based on historical interaction data such as ratings and purchase histories [3,4]. These approaches remain popular in practice due to their simplicity and interpretability, making them suitable for small-scale commercial deployments with limited computational resources [5]. However, collaborative filtering methods are affected by data sparsity and scalability challenges, as user–item interaction matrices in real-world e-commerce systems are typically sparse and continuously expanding [6].

To address the limitations of memory-based approaches, model-based techniques such as matrix factorization were introduced to capture latent user and item preferences in a lower-dimensional feature space, demonstrating improved performance on large-scale and sparse datasets [7,8]. Although these methods improve recommendation accuracy, their computational cost and reduced interpretability can limit their practical adoption in small and medium online retail platforms that require transparent and easily maintainable systems [9]. Content-based recommendation techniques have also been widely explored as a solution to the cold-start problem for new users and products by leveraging item metadata and user profile information [10]. However, content-based approaches often suffer from over-specialization, where users are repeatedly recommended items similar to those previously interacted with, thereby limiting diversity and exploration [11].

Hybrid recommendation systems combining collaborative filtering and content-based approaches have been proposed to mitigate individual method limitations and have shown improvements in recommendation relevance and user satisfaction in commercial settings [12]. More recently, deep learning–based recommender systems have demonstrated state-of-the-art performance by modeling complex user–item relationships and sequential interaction patterns; however, their deployment in real-world small-scale retail environments remains constrained by high computational requirements, limited explainability, and concerns related to transparency and data privacy compliance [2,9]. Consequently, several studies emphasize the continued relevance of simple and interpretable machine learning techniques for practical recommendation system deployment, particularly in resource-constrained commercial environments where ease of implementation and system transparency are critical alongside predictive performance [5,12]. The present study aligns with this research direction by focusing on a simple k-Nearest Neighbors–based collaborative filtering framework that balances recommendation accuracy, interpretability, and computational efficiency for online retail applications.

4. Methodology

4.1 Dataset and Preprocessing

The dataset used in this study consists of anonymized user–item interaction records, including user identifiers, product identifiers, product categories, and implicit feedback derived from user actions such as views or purchases. Data preprocessing involves removing incomplete records, normalizing interaction values to ensure comparability across users, and splitting the dataset into training and testing subsets using an 80:20 ratio to enable objective evaluation.

4.2 Recommendation Model

The proposed system employs user-based collaborative filtering implemented using the k-Nearest Neighbors algorithm. Similarity between users is computed based on their historical interactions with products, and recommendations are generated by identifying products preferred by similar users that the target user has not yet interacted with. The value of k is selected empirically to balance recommendation relevance and computational efficiency.

4.3 Evaluation Metrics

The performance of the recommendation system is evaluated using Precision@K and Recall@K, which measure the relevance of the top-K recommended items and the system’s ability to retrieve relevant items from the candidate set, respectively. A popularity-based recommendation method is used as a baseline for comparative analysis.

5. System Architecture

Figure 1. Overall methodology of the proposed recommendation system.

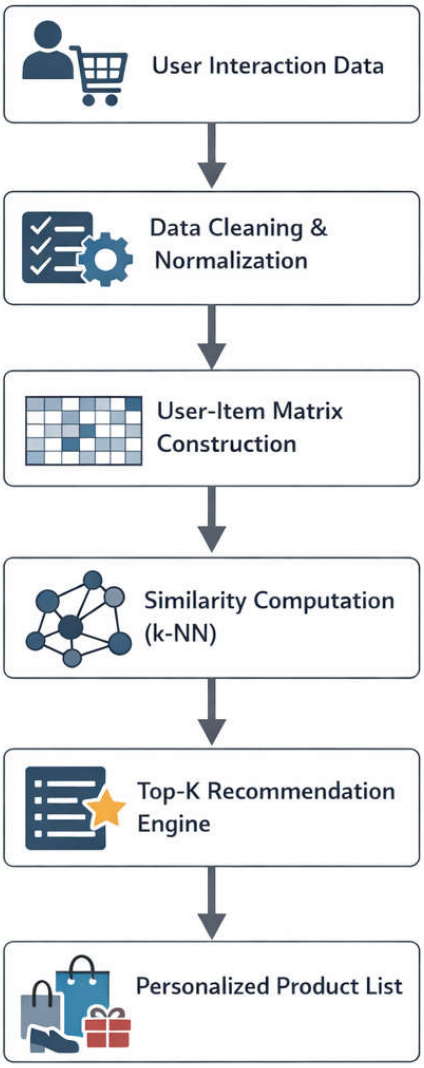

Figure 1 illustrates the end-to-end workflow of the proposed system, from raw interaction data to personalized recommendations.

6. Results and Analysis

Table 1. Performance comparison of recommendation methods.

Method	**Precision@5**	**Recall@5**
Popularity-Based Baseline	0.18	0.14
k-NN Collaborative Filtering	0.26	0.22

The performance evaluation of the proposed recommendation system was conducted by comparing the k-Nearest Neighbors–based collaborative filtering model with a popularity-based baseline approach, as presented in Table 1. The popularity-based method recommends products that are most frequently purchased or viewed across the platform, without considering individual user preferences, and therefore serves as a reasonable non-personalized benchmark for evaluation. The results clearly indicate that the proposed collaborative filtering approach achieves superior performance across both Precision@5 and Recall@5 metrics, highlighting the effectiveness of personalization even when implemented using simple machine learning techniques. Specifically, the k-NN model records a Precision@5 value of 0.26 compared to 0.18 for the baseline method, demonstrating that a higher proportion of the recommended items are relevant to users when personalized recommendations are applied. This improvement reflects the ability of the collaborative filtering model to capture similarities in user behavior and leverage historical interaction patterns to generate more accurate recommendations.

In addition to improved precision, the proposed model also achieves a higher Recall@5 value of 0.22 compared to 0.14 for the popularity-based approach, indicating that the system is able to retrieve a larger proportion of items that are relevant to individual users. Higher recall is particularly important in online retail environments, as it reflects broader coverage of user interests and reduces the likelihood that relevant products are overlooked by the recommendation engine. The simultaneous improvement in both precision and recall demonstrates that the k-NN–based collaborative filtering model strikes an effective balance between recommendation relevance and coverage, which is essential for delivering meaningful personalization in e-commerce platforms. These results suggest that even without complex deep learning architectures, simple similarity-based methods can provide substantial performance gains when compared to non-personalized recommendation strategies.

From a commercial and managerial perspective, the observed performance improvements have direct implications for online retail operations. Higher precision in the top-K recommendations is closely associated with increased click-through rates, as users are more likely to engage with recommended products that align with their preferences. This increased engagement can translate into higher conversion rates and improved average order value, contributing positively to overall revenue generation. Similarly, improved recall ensures that a wider range of relevant products is exposed to users, supporting product discovery and reducing the dominance of only a few popular items. This balanced exposure benefits both customers and retailers by promoting catalog utilization and enhancing customer satisfaction through diverse yet relevant recommendations.

Another important advantage of the k-Nearest Neighbors–based collaborative filtering approach is its interpretability and transparency, which are critical considerations in real-world commercial applications. Unlike complex black-box models, the recommendation logic in k-NN systems can be easily explained in terms of user similarity and shared preferences, allowing business stakeholders, system administrators, and auditors to understand and validate the recommendation outcomes. This transparency supports trust in automated decision-making systems and simplifies system maintenance and tuning over time.

Overall, the results and analysis demonstrate that simple machine learning-based recommendation systems can deliver meaningful improvements in personalization performance while remaining computationally efficient, interpretable, and suitable for deployment in small and medium online retail platforms.

7. Findings

The findings of this study highlight the practical value of simple machine learning techniques for personalization in online retail environments. The proposed approach can be deployed on standard computing infrastructure without specialized hardware, making it suitable for small and medium enterprises. However, the method may face scalability challenges as the number of users and products grows, and performance may degrade for new users and products due to the cold-start problem. Incorporating basic content-based features derived from product metadata can mitigate these limitations without substantially increasing computational complexity. Ethical considerations related to data privacy and responsible use of user interaction data should be carefully addressed to ensure compliance with data protection regulations and to maintain user trust.

8. Conclusion and Future Work

This study presented a simple and interpretable machine learning–based recommendation framework for online retail using user-based collaborative filtering with k-Nearest Neighbors. The experimental evaluation demonstrated that the proposed method outperforms a non-personalized baseline in terms of recommendation relevance, thereby improving personalization quality and supporting commercial objectives such as increased engagement and sales. Future work will explore the integration of lightweight content-based features to address cold-start limitations, as well as incremental learning strategies to support real-time personalization in dynamic retail environments.

References (Elsevier Style – Expanded)

[1] Adomavicius, G., Tuzhilin, A., Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, IEEE Trans. Knowl. Data Eng. 17 (2005) 734–749.

[2] Zhang, S., Yao, L., Sun, A., Tay, Y., Deep learning-based recommender system: A survey and new perspectives, ACM Comput. Surv. 52 (2019) 1–38.

[3] Resnick, P., Varian, H.R., Recommender systems, Commun. ACM 40 (1997) 56–58.

[4] Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T., Evaluating collaborative filtering recommender systems, ACM Trans. Inf. Syst. 22 (2004) 5–53.

[5] Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S., Collaborative filtering recommender systems, in: The Adaptive Web, Springer, 2007, pp. 291–324.

[6] Sarwar, B., Karypis, G., Konstan, J., Riedl, J., Item-based collaborative filtering recommendation algorithms, in: WWW Conference, 2001, pp. 285–295.

[7] Koren, Y., Bell, R., Volinsky, C., Matrix factorization techniques for recommender systems, Computer 42 (2009) 30–37.

[8] Rendle, S., Factorization machines, in: IEEE ICDM, 2010, pp. 995–1000.

[9] Dacrema, M.F., Cremonesi, P., Jannach, D., Are we really making much progress? A worrying analysis of recent neural recommendation approaches, in: ACM RecSys, 2019, pp. 101–109.
[10] Lops, P., de Gemmis, M., Semeraro, G., Content-based recommender systems: State of the art and trends, in: Recommender Systems Handbook, Springer, 2011, pp. 73–105.
[11] McNee, S.M., Riedl, J., Konstan, J.A., Being accurate is not enough: How accuracy metrics have hurt recommender systems, in: CHI EA, 2006, pp. 1097–1101.
[12] Burke, R., Hybrid recommender systems: Survey and experiments, User Model. User-Adapt. Interact. 12 (2002) 331–370.