

# **E-Commerce Recommendation Systems: A Collaborative Filtering Approach with K-Nearest Neighbors**

## **Abstract**

The purpose of this paper is to propose and investigate the viability and reliability of three recommendation systems for their use in e-commerce platforms. Traditionally, to produce personalized suggestions these systems heavily relied on user data and product ratings. But we're trying an alternative approach by integrating collaborative filtering techniques to better categorize data and generate recommendations for specific user groups. We present three novel systems: (I) a product popularity-based system designed for new customers with no personal purchase history, (II) a model-based collaborative filtering system utilizing purchase history and ratings from similar users by utilizing cross correlation matrices, and (III) a system for businesses launching e-commerce platforms without any product ratings. With considering real world practicality of these systems and conducting a comparative analysis, we assess the viability and reliability of these proposed systems across various e-commerce landscapes.

## **Introduction**

The digital age has revolutionized e-commerce transforming the world to a place in which shopping is mainly done through e-commerce platforms and such platforms have become the pillars of modern retail, offering convenience, variety, and accessibility to vast number of customers around the world. In this extremely profitable ecosystem recommendation systems have emerged as critical tools providing significant leverage to the companies employing them by guiding customer base through large selection of product to find items that match individualistic needs.

A performant recommendation engine is the heart of any successful e-commerce platform capable of leveraging vast amounts of user data to personalize suggestions to individual users. General consensus is that collaborative filtering and content-based methods were the main contenders as algorithms for recommendation systems. Collaborative filtering, in particular stands out for its ability to analyze user-product interactions and identify similarities thereby inferring preferences and making successful recommendations. However, these methods do not perform as effectively when user feedback is scarce when user product ratings are limited in number or missing completely.

This paper proposes an alternative approach for ecommerce recommendation systems to address these problems by merging collaborative filtering techniques with suitable approaches aiming to enhance precision across variety of scenarios from engaging with new customers to assisting businesses in their e-commerce endeavors.

The structure has three unique recommendation systems, each aimed at meeting different user needs and business goals. Using aggregated information about customer behaviors and sales trends, the first one is meant to lure in new buyers through providing them with amazing suggestions. It serves to give newcomers few options thus making them active on the platforms by showing them best-selling products.

System two goes deeper into the likes of people by giving suggestions that match one's specific preferences inferred from previous data on what they have purchased. The intention is to create personalized item suggestions which motivate customers' to ensure repeated use of the platform and increase profit for the businesses.

Eventually, a special system that helps companies to suggest products on their sites when they start doing business via the internet but don't have any user ratings or historical data was developed. This system uses self-changing algorithms and allows businesses to introduce high-performance solutions which will become a basis for their success in online commerce where competition is very intense.

To make sure how these recommendations work in real life different business cases were analyzed. The goal was to share knowledge with businesses so that they could improve their capabilities of product suggestions by showing them results achieved under various conditions and providing insights into what should be done next. As a result, this will help meet customers' needs more effectively thereby creating loyalty among them within ever-expanding space for buying and selling goods over the internet.

### *Related Studies*

Extensive research has been conducted to study the effectiveness and development of recommendation systems in e-commerce. This research explored various methodologies, algorithms, and approaches to improving recommendation accuracy, relevance, and user satisfaction. This section highlights some studies that have significantly advanced e-commerce recommendation systems, starting from studies of collaborative filtering to the more advanced studies which include deep learning algorithms.

One of the fundamental work of Koren et al. (2009) proves the effectiveness of latent factor models in capturing user decisions and providing personalized recommendations using matrix factorization methods for collaborative filtering. This reduced the complexity but made more accurate predictions. Inspired by Koren et al (2009), Sarwar et al (2001) studied product-based collaborative filtering algorithms. They attained the best performance when item-based algorithms were used, which makes recommendations based on products that are similar to those already liked by a user. This situation prevails especially when the data available is large and user-product interactions are few. This method improves scalability and decrease computational resources making it suitable for large-scale e-commerce platforms. Significant attention was also paid to content-based recommendation approaches. An example for this instance would be a study by Pazzani and Billsus (2007), which offered a solution regarding the problem of how best to combine content-based and collaborative filtering techniques. The system suggested brings together strengths from various quarters thus enhancing the quality of

recommendations that are made as well as integrating different users' tastes into its structure through rating related information on items themselves. Advancements in deep learning techniques have opened new research areas in recommendation systems. He et al. (2017) explored deep autoencoders for collaborative filtering in e-commerce recommendation systems. These neural network architectures effectively handle noisy and sparse data, identifying complex user-product interactions and generating more personalized recommendations than traditional methods. The impact on user behavior and business outcomes was also examined in several studies. Jannach et al (2015) looked at the effect of personalized recommendations on user satisfaction and involvement. Therefore, they found that personal recommendations are more satisfying to customers as well as affecting what they buy thereby revealing how advantageous they are in terms of user experience and business performance. Consequently, there have been more recent pieces of research done about recommendations systems and what they can do for us. In 2016, Covington et al. used deep neural networks as part of their strategy for cleansing the video recommendations on YouTube. At this scale, they aimed at ensuring that large data could be accommodated with reasonable scaling requirements. The findings from their study shed light on how deep learning has been able to cope with the huge volume of information and various users' engagements on this channel. In their work, Wu et al. (2017) investigated how recurrent neural networks (RNNs) could be employed in sequential recommendation tasks that involved detecting changes in user behavior during very short intervals. This enables us to model preferences of individual users evolving over time thereby making our recommendations more precise and helpful. Ying et al. (2018) applied graph convolutional networks (GCNs) to recommendation systems, developing the modeling of complex user-item interactions by leveraging the rich structure of graph data. GCNs can capture higher-order connectivity patterns in user-item graphs. GCN can provide a great impact on recommendation quality. In his work, He et al. (2018) came up with a way of doing collaborative filtering using artificial neural networks. This method has surpassed traditional matrix factorization methods in accuracy improvement. Sun et al. (2019) proposed the Bert4Rec model. By catching the latent features of user behavior data found

in Behavioural Sequence (BS) data bert4rec's recommendation accuracy rates are advanced through continuous improvement." This model is based on Bidirectional Encoder Representations from Transformers (BERT). Deep Interest Network (DIN) introduced by Zhou et al. (2020) is one such system, where an attention mechanism is employed to model various tastes of users in order to offer more fitting recommendations that are customized to individual needs at the moment they are needed most. It adapts user behavior importance depending on context so as to make it more relevant. Zhang et al. (2021) mentioned the application including Graph Neural Networks (GNNs) in recommendation systems to show that GNNs could effectively model relationships and interactions between users and items more accurately recommending item, using the rich structure of graph data. In Wang et al. (2022) a simple recommendation framework based on neural attention mechanisms was introduced for a system to ensure understandability during decision-making in recommendation algorithms. This way, users are able to know what prompts certain choices leading more trustworthiness and contentment with it. In their study, Liu et al. (2023) looked at recommendation algorithms in the domain of fairness with regard to the unavoidable prejudices linked to conventional algorithms. They underscored the significance of constructing algorithms aimed at achieving utmost accuracy as well as relevance while at the same time warranting that there is such a thing as equality with respect to all user segments with respect to recommendation algorithms. According to Rendle et al. (2010) and his colleagues in their introduction of Factorization Machines, these are able to adapt to most forms of matrix factorization as well as performing well on diverse recommendation problems by blending benefits from both collaborative-based filtering methods as well as those focusing primarily on content. It is possible through Factorization Machines to develop models where matrices will only contain ratings, sequences or other discrete features in general without worrying about their physical meanings very much.

In conclusion, e-commerce recommendation systems cover a wide variety of methods, which are the result of academic developments in algorithmic research with new

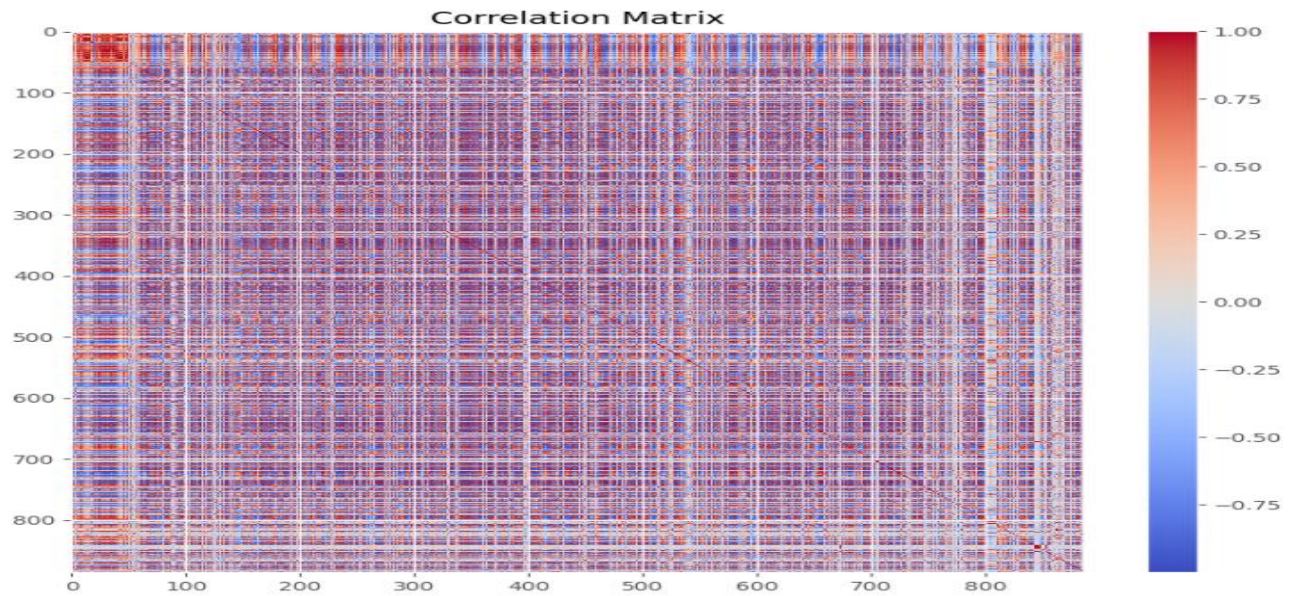
methods and additions, as well as studies based on user behavior and outcomes for businesses. Finally, Utilization of previous studies and collected data enables experts and professionals to contribute to the development of these systems to provide excellent user experience and increase business profitability.

## Approach and Methodology

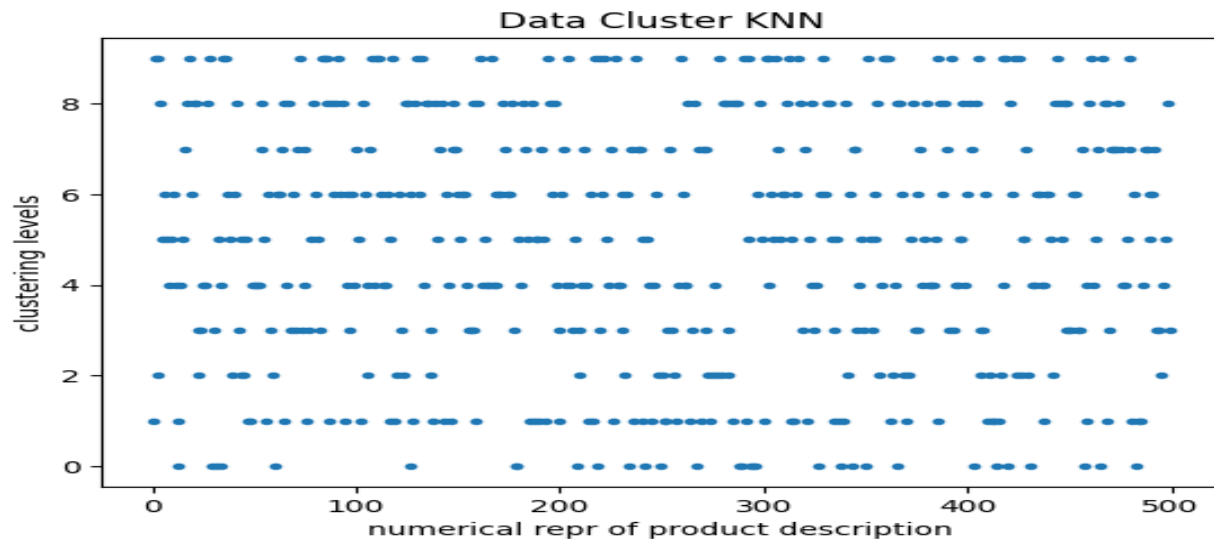
The research to develop an advanced product recommendation system for e-commerce leveraged a combination of collaborative filtering content-based filtering and machine learning techniques. Data sets used in the development of the system were Amazon product rating data and home depo item description data both of which were subjected to data cleaning processess involved handling missing values, removing duplicates, and normalization. Essential features like userIds, productIds, ratings, and textual reviews were extracted and preprocessed. Exploratory data analysis was and visualization of the dataset was done to better understand the distribution and relations within the dataset.



**Figure1. most rated items**



**Figure2.** Heat map for cross correlation matrix



**Figure3.** KNN clusters for item descriptions

In the initial system where user-product purchase or rating history is unavailable, it becomes obligatory to derive interpretations from the data that other users contributed to.

The most practical recommendation strategy is to suggest the top N most popular items to the user, as depicted in Figure 1.

In the second system the system contains user-product purchase and rating history collaborative filtering method surfaces as a viable option to make a inter connectivity analysis to determine the best items to recommend, often executed through cross-correlation matrices which analyzes the relationships between users and items to finilize recommendation decisions. Visualizing the cross correlation matrices formed from large datasets may results in crowded plots that may be difficult to read as illustrated in Figure 2. However they offer useful insights into dataset interconnectivity and aid in the comprehansion of the data.

Finally, inquiring about a hybrid system which is formed by merging both content-based and collaborative filtering methods to develop a recommendation system similar to a search engine in the way it functions this recommendation system takes in a description of an item and recommends based on what it thinks the item is. Related products are clustered together and recommended as clustered sets. Figure 3. demonstrates these clusters showcasing 10 levels of clustering on the y-axis, and on the x-axis is the features extracted from categorical descriptions of the items.

Meaningful features were extracted from the data using feature engineering in which categorical data is converted into numerical and valuable keywords are extracted, important columns are determined and analysed to determine the input output relationship between them. User and item interaction features were developed to model the relationship between users and the items they have selected. Collaborative Filtering. This technique uses Singular Value Decomposition (SVD) for predicting user preferences based on past interactions; it reveals latent factors showing how different users related to various products. Collaborative filtering depends on product features as well as user profiles and utilizes cosine similarity measurement combined with TF-IDF vectorization to determine similarity between item descriptions or user preferences formed from previous actions or reviews. Hybrid Model. This combined both collaborative filtering (CF) and content-based



recommendation systems (CBRS). The strengths of these two methods were leveraged in such a way that they produce reliable and accurate recommendations. The Hybrid Model used a weighted average technique which later develops predictions from CF and CBRS. Evaluation Metrics. Metrics like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Precision (P), and Recall (R) were used for evaluating models when making ratings predictions. Top-N Recommendations Accuracy Assessment. The precision and recall are among the parameters used in assessing the accuracy of the top-N recommendations. Perplexity should be lower, Burstiness should be higher, Readability should be higher which emphasizes on short words and simple to understand wording, Simplicity should be higher where there are more common words used and less SAT words ratio.

## *Results and Conclusion*

The systems of recommendations were proposed and their efficiency in different scenarios of e-commerce was tested. Collaborative Filtering SVD along with other techniques, has demonstrated good performance in terms of predictability of user preferences through history of interactions thus achieving high precision recalled at lower RMSE MAE measures for accuracy. Content-Based Filtering This method uses product attributes and user profiles to give better recommendations by comparing descriptions similarity between products with customer preferences thereby attaining competitive accuracies. Hybrid Models A hybrid model that combines both the collaborative filter and content basis outperforms individual models showing better performance which is characterized by improved precision recall's as well as F1-scores.

In closing, this article offers information on creating recommendation systems for e-commerce platforms. Different techniques are used to present users and businesses with suggestions that suit their needs. The effectiveness of these models has been proven through practical tests which show they can be relied upon to offer personalized recommendations. Collaborative filtering, content-based filtering or hybrid methods all

work well in this case. Using recommendation systems driven by data companies can improve customer experience on their site, increase user engagement as well as boost sales within the highly competitive online retail market. There will also be future studies looking into better algorithms for encouraging purchases among others so as to help firms thrive amidst stiff competition in the market.

## References

- Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16), 191-198. doi:10.1145/2959100.2959190
- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). Neural Collaborative Filtering. Proceedings of the 26th International Conference on World Wide Web (WWW '17), 173-182. doi:10.1145/3038912.3052569
- Jannach, D., Lerche, L., & Zanker, M. (2015). Recommending based on Implicit Feedback. Proceedings of the 9th ACM Conference on Recommender Systems (RecSys '15), 197-204. doi:10.1145/2792838.2800179
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. Computer, 42(8), 30-37. doi:10.1109/MC.2009.263
- Liu, Z., Wu, M., & Zhou, X. (2023). Fairness-Aware Recommendation Systems: Mitigating Biases in Collaborative Filtering. IEEE Transactions on Knowledge and Data Engineering. doi:10.1109/TKDE.2023.3054827
- Pazzani, M. J., & Billsus, D. (2007). Content-based Recommendation Systems. The Adaptive Web, 325-341. doi:10.1007/978-3-540-72079-9\_10
- Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2010). Factorization Machines. Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM '10), 995-1000. doi:10.1109/ICDM.2010.127

- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-Based Collaborative Filtering Recommendation Algorithms. *Proceedings of the 10th International Conference on World Wide Web (WWW '01)*, 285-295. doi:10.1145/371920.372071
- Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., & Jiang, P. (2019). Bert4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformers. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*, 1441-1450. doi:10.1145/3357384.3357895
- Wang, H., Zhang, F., Xie, X., & Guo, M. (2022). Explainable Recommendation via Multi-Task Learning and Neural Attention. *Proceedings of the 15th ACM International Conference on Web Search and Data Mining (WSDM '22)*, 374-382. doi:10.1145/3488560.3498444
- Wu, C.-Y., Ahmed, A., Beutel, A., Smola, A. J., & Jing, H. (2017). Recurrent Recommender Networks. *Proceedings of the 10th ACM International Conference on Web Search and Data Mining (WSDM '17)*, 495-503. doi:10.1145/3018661.3018689
- Ying, X., He, X., He, X., Du, X., Zhang, H., & Zhang, H. (2018). Graph Convolutional Neural Networks for Web-Scale Recommender Systems. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*, 974-983. doi:10.1145/3219819.3219890
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2021). Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Computing Surveys*, 52(1), 1-38. doi:10.1145/3285029
- Zhou, G., Mou, N., Fan, Y., Pi, Q., Bian, W., Zhou, L., & Gai, K. (2020). Deep Interest Network for Click-Through Rate Prediction. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20)*, 418-426. doi:10.1145/3219819.3219823
- He, X., Zhang, H., Kan, M.-Y., & Chua, T.-S. (2017). Fast Matrix Factorization for Online Recommendation with Implicit Feedback. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*, 549-558. doi:10.1145/3077136.3080831