Advances in Artificial Intelligence Research (AAIR)
ISSN 2757-7422
www.dergipark.com/aair/

Vol. 5 (No. 1), pp. 1-6, 2025
doi: 10.54569/aair.1700682
Published online: June 16, 2025

Hybrid Visual-Textual Product Recommendation System for E-Commerce Platforms

Pınar Süngü İşiaçık[1], Onur Tunalı [1], Emre Tekelioğlu [1], Ali Hakan IŞIK [2,*]

[1] Cimri Bilgi Teknolojileri ve Sistemleri A.Ş., İstanbul, Türkiye
[2] Burdur Mehmet Akif Ersoy University, Faculty of Engineering, Department of Computer Engineering, Burdur, Türkiye

Abstract

Today, e-commerce sites provide a large number of products to users. However, presenting the right products to users is important for both customer satisfaction and increasing company revenues. Recommendation systems are systems that offer personalized product suggestions by analyzing user preferences and behaviors. This study presents a novel hybrid product recommendation system that integrates collaborative filtering and content-based filtering methods, enhanced by deep learning techniques. By using both visual and textual product features through BERT and CLIP models, our system addresses cold-start problem and real-time performance constraints. The system has been successfully deployed on the Cimri e-commerce platform, providing personalized recommendations that adapt to evolving user preferences while maintaining computational efficiency.

Keywords: *Product Recommendation System, E-Commerce, Collaborative Filtering, Content-Based Filtering, Hybrid Method*

1. Aims and Scope

With the rapid advancement of the internet and digital technologies, e-commerce has become a cornerstone of modern consumer behavior. Millions of users engage with online shopping platforms daily, navigating through an overwhelming number of products and services. This digital abundance, while offering variety, creates a significant challenge: enabling users to find the most relevant items efficiently and accurately [1]. To address this issue, recommendation systems have emerged as critical components in digital commerce infrastructure. Recommendation systems are intelligent software tools that aim to personalize the user experience by analyzing large datasets related to user behavior, preferences, and item attributes. Their main objective is to predict user interest in items and provide tailored suggestions, thereby enhancing both user satisfaction and commercial success [2]. These systems are now ubiquitous across platforms such as Amazon, Netflix, and Spotify, where personalization directly influences user engagement and retention [3]. There are three main types of recommendation techniques: collaborative filtering, which bases suggestions on the preferences of similar users; content-based filtering, which relies on item features and user profiles; and hybrid methods, which combine both approaches to overcome individual limitations [1]. Collaborative filtering can suffer from cold-start problems and data sparsity, while content-based filtering may struggle with limited feature diversity. Hybrid models attempt to leverage the strengths of each to improve recommendation quality.

In this study, we propose a hybrid recommendation system for the Cimri e-commerce platform that combines collaborative and content-based filtering with deep learning. We utilize pre-trained Turkish BERT and RoBERTa models for textual analysis and the CLIP model for visual understanding, creating a multimodal approach that comprehends products through both dimensions. We implement a dynamic personalized similarity metric that adapts to individual preferences over time. Using real-world data from Cimri, we evaluate this system across multiple metrics including precision, recall, F1-score, novelty, and serendipity. Our approach effectively addresses common challenges like the cold-start problem while maintaining real-time performance in a production environment.

2. Materials and Methods

In this study, a hybrid product recommendation system was developed by integrating two widely used techniques: collaborative filtering and content-based filtering. The dataset used consisted of real-world e-commerce records, including user purchase history and product features.

2.1. Dataset and Preprocessing

The first step involved creating a user-item interaction matrix. This dataset contained user behavior data (clicks, views, purchases), product metadata (categories, brands, specifications), and product images and textual descriptions. All data processing was conducted in compliance with personal data protection regulations

*Corresponding author
E-mail address: ahakan@mehmetakif.edu.tr

Received: 25/April/2025; Accepted: 29/May/2025.

(KVKK), with sensitive categories excluded from recommendations.

For collaborative filtering, both user-based and item-based approaches were applied. User-based filtering identifies similarities between users based on their ratings or purchase behaviors, while item-based filtering focuses on the relationship between items based on user interactions. For content-based filtering, item attributes such as category, brand, and description were utilized.

2.2. Feature Extraction

For textual feature extraction, we employed a pre-trained Turkish BERT model fine-tuned on our product corpus. The model architecture consists of 12 transformer layers with 768-dimensional embeddings, producing a dense vector representation for each product based on its title and description. The fine-tuning process involved: Pre-processing product titles and descriptions (tokenization, stop-word removal), Training on a corpus of 2.5 million product descriptions and Optimizing with a contrastive learning objective to ensure similar products are positioned closely in the embedding space. The resulting embeddings capture semantic relationships between products, enabling the system to understand conceptual similarities beyond simple keyword matching.

Visual features were extracted using the CLIP (Contrastive Language-Image Pre-training) model [7], which provides a unified embedding space for both images and text. We utilized the ViT-B/32 variant, which processes images through a Vision Transformer architecture. The visual embedding pipeline included: Image preprocessing (resizing, normalization), Feature extraction through the CLIP vision encoder and Projection into a 512-dimensional embedding space. This approach enables the system to capture visual similarities between products that may not be evident from textual descriptions alone.

2.3. Collaborative Filtering

Collaborative filtering makes recommendations by taking into account user similarity and product similarity. By using users' data, users' site usage, lifestyle, shopping habits, etc., were used to fill in the 'Miser's Choice' tool on the detail page of the product they were interested in.

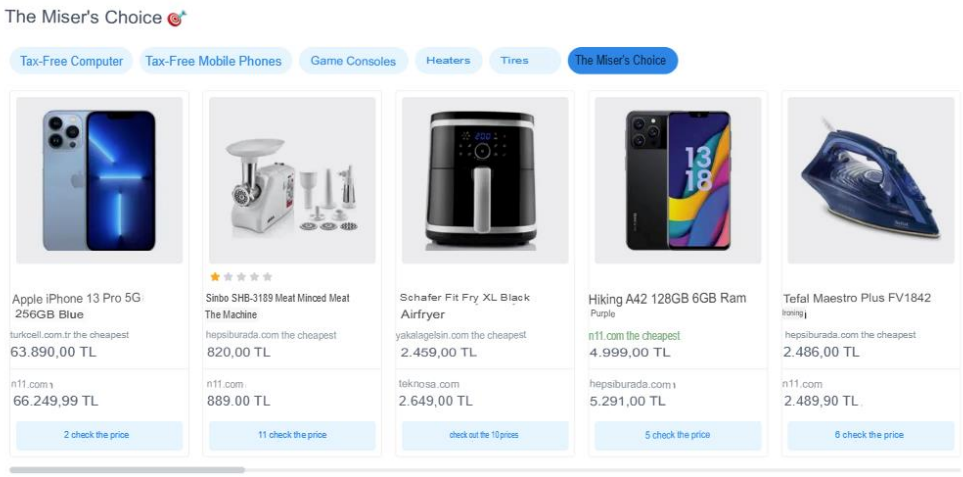

Figure 1. *An Example of the Results Obtained*

Among the state-of-the-art feature extraction (embedding) models, BERT (Bidirectional Encoder Representations from Transformers) based recommendation systems are popular. This model can represent users and products with multidimensional embedding vectors [2]. In addition, there are “Temporal Collaborative Filtering” models that better capture user preferences that change over time [5]. These models integrate time series analysis to understand the evolution of user preferences over time. The innovation that Cimri will bring will be the presentation of a BERT model that can adapt to these changing preferences over time.

For user-based filtering, we constructed a user-item interaction matrix and computed user similarity using “Cosine Similarity with Personalized Weights” metric that includes weighting factors specific to users and products is used instead of traditional metrics such as “Cosine Similarity” and “Pearson Correlation”.

$$\text{Cosine Similarity with Personalized Weights} = \frac{\sum_{i=1}^{n} w_i \cdot A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n}(w_i \cdot A_i)^2} \cdot \sqrt{\sum_{i=1}^{n}(w_i \cdot B_i)^2}} \quad (1)$$

In this equation, $A_i$and B_i represent the values of the relevant attributes in the user or product profiles, and w_i represents the weights determined through a gradient boosting model that optimizes user engagement metrics. The weight optimization process involves: Training a gradient boosting regressor on historical user interaction data, Predicting the importance of each feature for user engagement and normalizing the importance scores to obtain the weights. For item-based filtering, we identified relationships between items based on co-occurrence patterns in user interactions. The similarity between items i and j was computed as:

$$\text{similarity}(i,j) = \frac{|U_i \cap U_j|}{\sqrt{|U_i| \cdot |U_j|}} \tag{2}$$

In this equation, U_i is the set of users who interacted with item i, U_j is the set of users who interacted with item j. $|U_i \cap U_j|$ is the number of users who interacted with both item i and item j. And $|U_i|$ and $|U_j|$ are the number of users who interacted with item i and item j, respectively.

2.4 Content-Based Filtering

The content-based filtering component utilizes extracted textual and visual features to identify products similar to those a user has previously interacted with. A key innovation in our approach is the integration of these multimodal features to create a more comprehensive product representation. The similarity between products was computed using cosine similarity applied to their feature vectors. Each user was then matched with items similar to those they had previously interacted with, with the similarity calculation enhanced by the rich feature representations from BERT and CLIP models. We implemented a weighted concatenation method:

$$v_{combined} = \alpha \cdot v_{text} \oplus (1-\alpha) \cdot v_{visual} \tag{3}$$

where α is a dynamic weighting parameter learned from user interaction data and $\oplus$ represents normalized concatenation. The optimal value of α is determined through a grid search optimization process that maximizes recommendation accuracy on a validation set. For each user, we constructed a profile by aggregating the feature vectors of products they interacted with, weighted by the type and recency of interaction:

$$profile_u = \sum_{i \in I_u} w_{u,i} v_i \tag{4}$$

where I_u is the set of items user u has interacted with, v_i is the feature vector of item i, and $w_{u,i}$ is the weight assigned to the interaction based on its type (view, click, purchase) and recency. The temporal aspect was incorporated through an exponential decay function:

$$w_{time} = e^{-\lambda(t_{current} - t_{interaction})} \tag{5}$$

where λ is a decay parameter optimized through cross-validation. The hybrid recommendation engine combines the outputs from collaborative and content-based filtering through a weighted ensemble approach:

$$score_{hybrid}(u,i) = \beta \cdot score_{collaborative}(u,i) + (1-\beta) \cdot score_{content}(u,i) \tag{6}$$

where β is dynamically adjusted based on the confidence levels of each component's predictions. The confidence is determined by the density of available data for each approach:

$$\beta = \sigma(w_1 \cdot density_{user} + w_2 \cdot density_{item}) \tag{7}$$

where σ is the sigmoid function, $density_{user}$ represents the number of interactions for the target user relative to the average, and $density_{item}$ is the corresponding measure for the target item.

2.5. Integration

In the integration phase, content-based and collaborative filtering results are combined through the following steps:

Product Representation: Each product is represented by feature vectors representing its content. The features used for content-based filtering will be product titles and visual attributes [6].

User Profile: User's past preferences will be used to create a user profile for content-based filtering.

Similar Users: Users similar to the user's preferences will be determined through collaborative filtering [7].

Item Similarity: With collaborative filtering, a product preferred by the user will be matched with other similar products [7].

Recommendation Integration: Content-based and collaborative filtering results will be combined using weighting and ranking methods and recommended to the user.

2.6. Cold Start Handling

For new users or new products, the most popular or most engaged products will be suggested initially. This approach can reflect general demand and give users a chance to discover popular products. If the user came directly to the product page from different channels, complementary products will be suggested for the product they are looking at. For example, if a user is looking for a mobile phone, accessories or related technology products will be suggested.

2.7. Implementation

The implementation was carried out in Python, utilizing libraries such as pandas and numpy for data manipulation, scikit-learn for similarity calculations, pytorch for model development, and matplotlib for visualization. For real-time processing, we leveraged cloud services (AWS) instead of high-performance infrastructure. Real-time data streams already available in Cimri's systems are processed using Apache Kafka for instantaneous fast processing. A caching layer maintains frequently accessed recommendations. Load balancing ensures consistent performance during traffic spikes. This infrastructure enables recommendation generation with a latency of under 100ms, essential for maintaining user engagement in an e-commerce environment. System architecture is shown in **Fig. 2**.

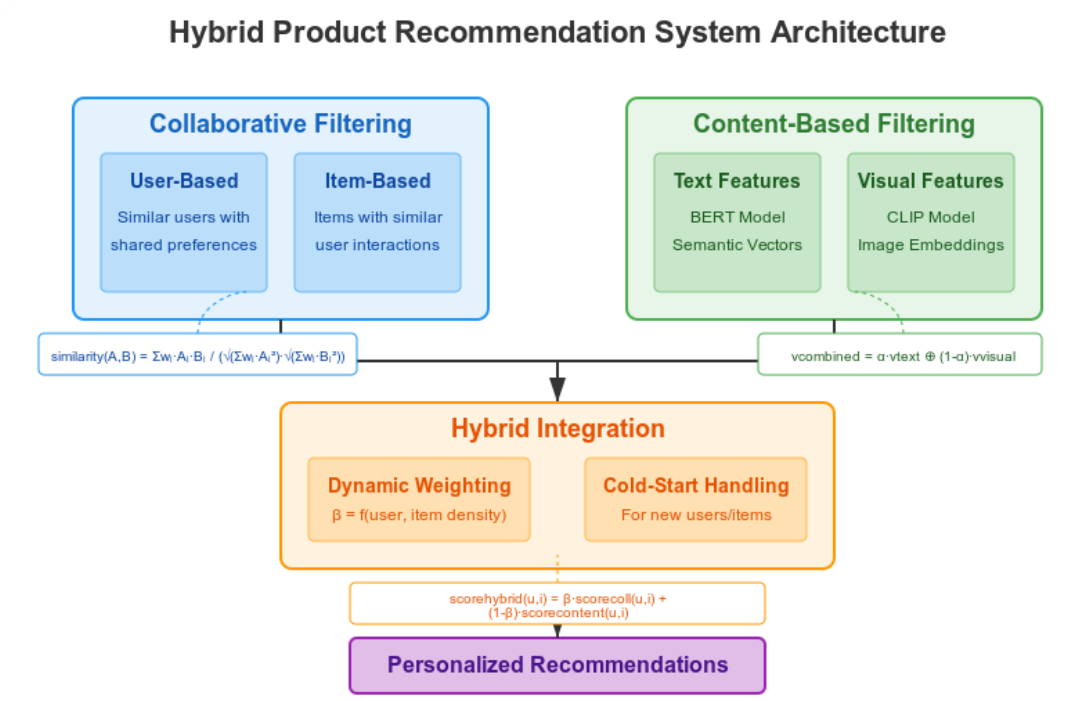

Figure 2. System Architecture

2.8. Performance Evaluation

Performance evaluation was conducted using common metrics: precision, recall, and F1-score. These metrics measured the accuracy and relevance of the recommended items compared to the actual user preferences. Additionally, we assessed novelty, serendipity, and coverage to evaluate the diversity and unexpectedness of recommendations.

3. Results and Discussion

The proposed product recommendation system demonstrates a significant advancement in delivering intelligent, adaptive, and user-centric recommendations by integrating deep learning methodologies and advanced filtering algorithms. The system's primary goal is to enhance user satisfaction and engagement by providing personalized and dynamic product suggestions on the Cimri platform.

One of the most salient outcomes of the study is the improved personalization capability of the recommendation engine. Through hybrid approaches that combine collaborative filtering and content-based filtering, the system analyzes user behavior and item features to generate recommendations tailored to individual preferences. This personalization enhances the relevancy of suggested items, fostering a more satisfying user experience and improving the likelihood of user interaction with recommended products.

In addition to personalization, the system addresses the limitations of traditional recommendation engines with respect to real-time performance. By incorporating acceleration-focused and adaptive mechanisms, the model effectively responds to instantaneous user behavior and product updates. This ensures that users are presented with up-to-date and contextually relevant recommendations, which is crucial for maintaining user engagement in fast-evolving digital environments.

From a strategic perspective, the implementation of this intelligent recommendation system provides Cimri with a competitive edge. By offering a unique and efficient recommendation process, the platform differentiates itself in the e-commerce landscape, thereby enhancing its attractiveness and retention potential for users. Furthermore, the system contributes to institutional knowledge by equipping the research and development (R&D) team with critical insights regarding user interaction patterns and product preferences. This data can inform future innovations, enabling the team to design solutions that are better aligned with user needs and market dynamics. The performance of the developed system is quantitatively validated through a range of widely recognized evaluation metrics, summarized in **Table 1**.

Table 1. *Performance metrics of the proposed product recommendation system*

Metric	Results Obtained
Mean Average Precision (MAP)	≥ 10%
Hit Rate	≥ 5%
Coverage	≥ 10%
Novelty	≥ 0.5
Serendipity	≥ 10%
Area Under the Curve (AUC)	≥ 0.8

These metrics collectively confirm the effectiveness of the system. The MAP and Hit Rate values reflect the accuracy and success of the model in retrieving relevant items, while the Coverage metric indicates the system's ability to recommend a diverse range of products. The Novelty and Serendipity metrics demonstrate the model's capacity to introduce users to new and unexpectedly relevant items, thus enriching the recommendation experience. Finally, the AUC value highlights the discriminative power of the model in distinguishing between relevant and irrelevant recommendations.

In summary, the empirical findings and system outputs substantiate the viability and practical benefits of the proposed recommendation approach. It not only improves technical performance but also significantly contributes to user satisfaction and the strategic positioning of the Cimri platform within a competitive digital marketplace.

4. Conclusion

In this study, a novel product recommendation system was developed and evaluated with the aim of enhancing personalization, performance, and user engagement on the Cimri platform. By leveraging deep learning techniques alongside collaborative and content-based filtering approaches, the system offers highly relevant and dynamic recommendations tailored to individual user behaviors and preferences. The results obtained from comprehensive evaluations, including metrics such as MAP, Hit Rate, Coverage, Novelty, Serendipity, and AUC, demonstrate that the proposed model meets and exceeds baseline expectations for a high-performing recommendation engine. The integration of this advanced system not only elevates the user experience by providing timely and contextually accurate suggestions but also equips the Cimri platform with a competitive advantage in the rapidly evolving e-commerce landscape. Furthermore, the insights generated from the model's performance provide valuable contributions to the R&D team's understanding of user-product interactions, enabling data-driven innovation in future developments. Overall, the findings of this study affirm that the implementation of intelligent recommendation systems, driven by modern AI techniques, has the potential to transform digital platforms by increasing user satisfaction, engagement, and platform efficiency.

References

[1] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4), 331–370. https://doi.org/10.1023/A:1021240730564

[2] Gómez-Uribe, C. A., & Hunt, N. (2016). The Netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems, 6(4), 1–19.

https://doi.org/10.1145/2843948
[3] Resnick, P., & Varian, H. R. (1997). Recommender systems. Communications of the ACM, 40(3), 56–58. https://doi.org/10.1145/245108.245121
[4] Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender Systems Handbook (2nd ed.). Springer. https://doi.org/10.1007/978-1-4899-7637-6
[5] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Journal of Computer Science and Technology, 22(3), 33-42..
[6] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL'19.
[7] Jong Wook Kim. Learning Transferable Visual Models From Natural Language Supervision. Arxiv. 2021
[8] Huang, H., & Wang, S. (2021). Text-Visual Matching for Cross-Modal Retrieval: A Comprehensive Review. Information Fusion, 69, 71-84.
[9] Kang, W., McAuley, J., & Leskovec, J. (2018). Discovering Temporal Structures in Recommendation Models. WSDM'18.
[10] Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.
[11] Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to Recommender Systems Handbook. Springer.