Multimodal Hybrid Furniture Recommendation System Design

Project Overview

This project proposes a personalized furniture recommendation system for online retail platforms, aimed at enhancing user experience and purchase conversion among general e-commerce consumers. The system supports two recommendation modes:

Single-item recommendation: suggesting furniture pieces that align with a user's style preferences or current browsing interest;
Style-matching set recommendation: recommending compatible furniture items (e.g., sofa + coffee table + carpet) that form a visually harmonious combination.

To capture both user preferences and product characteristics, the system adopts a hybrid recommendation architecture, combining:

Content-based filtering, which extracts visual and textual features from product images and descriptions; and
Collaborative filtering, which leverages user-item interaction data (such as views, purchases, and ratings) to learn personalized preferences.

Visual style plays a central role in furniture selection. Therefore, this system incorporates deep visual embedding models (e.g., CNNs trained on labeled style data) to measure product compatibility and aesthetic coherence. For new users or products, the system falls back on visual and category-based similarity to address the cold-start problem.

The overall goal is to build a functional and scalable recommendation prototype using open-source datasets (e.g., DeepFurniture, Amazon reviews, Kaggle product data), and to evaluate its effectiveness via ranking metrics such as NDCG and user satisfaction simulations.

Scope

This project focuses on the design of a furniture recommendation system tailored for online shopping platforms. The target users are general e-commerce customers looking for aesthetically pleasing and functional furniture, either as individual pieces or as coordinated sets.

Domain: Online furniture retail (e.g., IKEA, Wayfair, Alibaba Home)

Target Users: Casual consumers browsing or purchasing furniture for home use

Recommendation Types:

I.Single-item recommendation — similar items based on visual and textual content

II.Set-based recommendation — furniture combinations that match in style and usage (e.g., bedroom, living room)

Interface Design:

I.Web or mobile interface

II.Card-style layout showing product images, price, and style tags

III.Optional mockup: “You may also like…” / “Style-matching sets for your room”

User Interactions:

I.Implicit: clicks, views, add-to-cart, purchases

II.Explicit: likes, ratings, saved/favorited items

Recommendation Dynamics:

I.Periodically updated as user interaction history grows

II.For new users: ask for preferred style via onboarding questionnaire

III.For new items: use visual content-based similarity to match users with interest

Business Relevance:

I.Encourages upselling and cross-selling through bundle recommendations

II.Increases user retention and engagement via personalized shopping experience

III.Potential monetization through promoted products and style-focused campaigns

Datasets

To build a hybrid recommendation system that incorporates visual aesthetics, content semantics, and user preferences, this project combines multiple open-source datasets:

4.1 Primary Data Sources

1.Amazon Product Review Dataset – Home & Kitchen (Furniture Subset)

Provides millions of user-item interactions (ratings, purchases, reviews) across categories, including furniture.

Contains: user_id, item_id, rating, reviewText, category, and metadata (e.g., title, brand, price).

Enables collaborative filtering via user-item matrix construction.

Source: Julian McAuley, UCSD ([1]).

2. Kaggle Furniture Image Dataset

Contains labeled images of various furniture types (e.g., sofa, table, chair).

Useful for training visual feature extractors (e.g., ResNet, EfficientNet).

Fields: image, label, color, material (optional).

Source: Uday Sankar Mukherjee, Kaggle ([2]).

3. DeepFurniture Dataset (via HuggingFace)

High-quality synthetic images of 24,000+ indoor scenes with 170,000+ furniture instances.

Each item annotated with 11 fine-grained style tags (e.g., modern, rustic, classical).

Enables style embedding learning and aesthetic compatibility training.

Source: Kujiale Research, HuggingFace ([3]).

4.2 Dataset Usage Summary

Dataset	**Used for**	**Modality**
Amazon Reviews	User-item interaction, metadata	Text, ratings
Kaggle Images	Visual embedding pretraining	Images
DeepFurniture	Style modeling, compatibility	Images + tags

4.3 Limitations

Amazon metadata may lack real product images; requires image-URL matching or synthetic replacement.

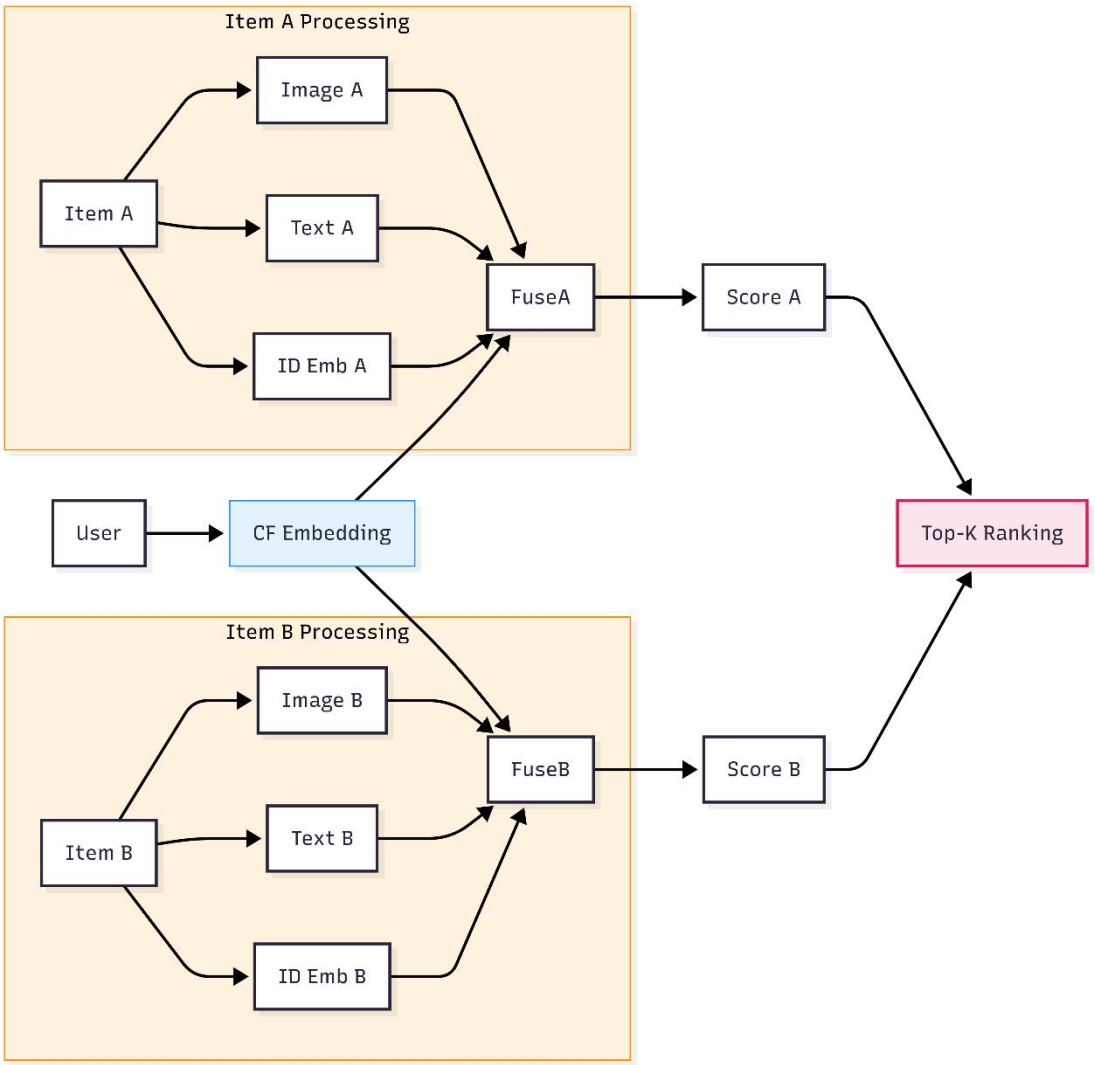
Kaggle image sets are small in scale and not fully labeled with style.

DeepFurniture contains synthetic renderings, which may not fully generalize to real e-commerce photos.

No single dataset contains both real user behavior and aligned product images with consistent style tags — manual alignment or weak supervision may be required.

Methods

This project adopts a hybrid recommendation framework that combines collaborative filtering (CF) and content-based filtering using multimodal features, including images, text, and product identity. The overall architecture is shown below,where each user–item pair is independently scored and ranked.



Figuare X

The system consists of three key components:

5.1 Model Architecture Overview

Our furniture recommender system adopts a hybrid architecture that integrates collaborative filtering (CF) with multimodal content features (image, text, and ID embeddings). For each user, the system processes candidate items independently via a

shared architecture that extracts item-specific features, fuses them with the user's latent embedding, and outputs a matching score. The final top-K recommendations are generated by ranking all item scores. Figure X illustrates the complete workflow for scoring and ranking two candidate items (A and B).

5.2 Fusion and Scoring

Each item is represented by a combination of three features:

Image features: Extracted from furniture images using a pre-trained CNN such as ResNet50, capturing visual style and shape:

$$\mathbf{v}_i = \mathrm{CNN}(x_i^{\mathrm{img}})$$

Text features: Encoded from product name and description using a BERT encoder to model semantics:

$$\mathbf{t}_i = \mathrm{BERT}(x_i^{\mathrm{text}})$$

ID embedding: A learnable dense vector that encodes item identity:

$$\mathbf{e}_i = \mathrm{Embedding}(i)$$

These embeddings are later concatenated for fusion.

5.3 Collaborative Filtering Embedding

To model personalized preferences, we extract user embeddings from historical interaction data (clicks, likes, purchases). Collaborative filtering[4] models such as Matrix Factorization or Neural CF can be used:

$$\mathbf{u}_u = f_{\mathrm{CF}}(u)$$

This embedding $\mathbf{u}_u$ encodes the user's latent interest and is combined with item features during scoring.

5.4 Fusion and Scoring

The image, text, ID, and user embeddings are concatenated and passed through a multi-layer perceptron (MLP) to compute a matching score:

$$\mathbf{z}_i = [\mathbf{v}_i \| \mathbf{t}_i \| \mathbf{e}_i \| \mathbf{u}_u]$$

$$\hat{r}_{ui} = \mathrm{MLP}_{\mathrm{fuse}}(\mathbf{z}_i)$$

Here, $\hat{r}_{ui}$ is the predicted relevance score between user u and item i. This flexible fusion enables the model to adaptively weigh different modalities.

5.5 Cold-start Strategy

Once scores are computed for all candidate items, the system ranks them and selects the top-K results for recommendation. For cold-start scenarios:

Cold-start users: use demographic info or onboarding preferences to initialize $\mathbf{u}_u$.

Cold-start items: can still be scored using image and text features

This design ensures robustness and flexibility in real-world applications.

Evaluation Plan

6.1 Evaluation Goals

The objective of this evaluation is to assess the effectiveness of the proposed hybrid recommendation system in both individual and set-based furniture recommendations. We aim to determine whether the system can:

Accurately match products to user preferences, both in terms of function and visual style;

Maintain ranking quality across various product categories;

Handle cold-start scenarios with minimal degradation in performance.

This evaluation ensures the system not only recommends relevant items but also maintains stylistic coherence and personalization across diverse user profiles.

6.2 Offline Evaluation Metrics

To quantitatively evaluate model performance, we will use standard offline metrics computed on a held-out test set:

Hit Rate@K: Measures the fraction of test interactions where the correct item appears in the top-K recommendations.

$$\text{Hit@K}=\frac{1}{|U|}\sum_{u\in U}\mathbb{I}\,(i^{*}\in \text{TopK}_{u})$$

NDCG@K (Normalized Discounted Cumulative Gain): Captures both relevance and ranking position:

$$\text{NDCG@K}=\frac{1}{|U|}\sum_{u\in U}\frac{1}{\text{IDCG}_{u}}\sum_{i\in \text{TopK}_{u}}\frac{2^{\text{rel}_{ui}}-1}{\log_{2}(\text{rank}_{i}+1)}$$

Precision@K / Recall@K: Evaluate the correctness and completeness of the top-K predictions.

Coverage: Measures the proportion of total catalog items that are ever recommended.

Intra-list Diversity: Computes diversity among recommended items based on content features (e.g., image embeddings or categories).

6.3 Cold Start Testing

Cold-start testing evaluates system robustness under limited data conditions:

Cold-start users: We simulate new users by hiding all their historical interactions. The system generates recommendations using only content-based features (e.g., style preference questionnaire or most similar user cluster).

Cold-start items: For newly added products with no interaction history, we evaluate whether the system can recommend them effectively using only image, text, and category features.

We will report changes in Hit@K and NDCG@K compared to regular users/items to quantify cold-start performance degradation.

6.4 User Study

To complement offline metrics, we propose a small-scale user study to qualitatively assess the relevance and aesthetic quality of the system's recommendations. A group of 5–10 student participants will interact with a simulated UI that displays:

Top-K recommendations for a given user profile or browsing history
Corresponding product images, titles, and style tags

Participants will rate each recommendation list based on:

I .Relevance to their preferences (1–5 scale)

II .Visual consistency of recommended sets (1–5 scale)

III .Overall satisfaction with the system output (1–5 scale)

The results will offer insight into user-perceived quality, especially in terms of style compatibility, which may not be fully captured by offline ranking metrics.

6.5 Computational Considerations

To ensure the system is feasible for real-world deployment or prototyping, we will monitor the following computational metrics during training and inference:

Model size: total parameter count (e.g., image encoder + fusion MLP)

Training time: time to convergence on sampled data

Inference latency: time to generate Top-K list per user

Storage cost: precomputed embedding storage if used

If scalability becomes an issue, we will explore:

Approximate nearest neighbor (ANN) search for content-based retrieval

Embedding caching and batch prediction to reduce real-time computation

These considerations will guide deployment decisions and model simplification trade-offs.

Conclusion & Feasibility

In this project, we proposed a hybrid furniture recommendation system that integrates content-based modeling and collaborative filtering. By combining visual features (from product images), textual semantics (from product descriptions), and ID embeddings, our model can better understand both the aesthetic and functional aspects of furniture. A user's preference is modeled through collaborative filtering embeddings, enabling personalized recommendations even with cold-start items via rich item-side features.

We adopted a modular architecture with a fusion-based scoring mechanism that incorporates multimodal item features and user embeddings. The system supports both single-item recommendations and style-coherent bundle suggestions, which are valuable in real-world furniture retail scenarios.

From a feasibility perspective:

We use open-source datasets such as [IKEA Furniture Dataset] and [Houzz Product Catalog], which contain multimodal content (images, texts, categories).

All feature extractors (ResNet for images, BERT for texts) are publicly available and pretrained, requiring no heavy training from scratch.

The hybrid model can be implemented using PyTorch or TensorFlow, and training can be conducted on a single GPU machine (e.g., RTX 3060/3090).

The recommendation pipeline can be evaluated offline using standard metrics such as Hit@K, NDCG@K, and Intra-list Diversity.

A lightweight user study can also be conducted to evaluate subjective style compatibility.

Overall, the project is technically feasible, educationally meaningful, and provides a practical introduction to multimodal personalized recommendation systems.

References

[1] McAuley, J., Pandey, R., & Leskovec, J. (2015). Inferring networks of substitutable and complementary products. *KDD*. [Amazon Review Data: http://jmcauley.ucsd.edu/data/amazon/]

[2] Mukherjee, U. (2022). Furniture Image Dataset. *Kaggle*. [https://www.kaggle.com/datasets/udaysankarmukherjee/furniture-image-dataset]

[3] Kujiale Research (2022). DeepFurniture Dataset. *HuggingFace Datasets*. [https://huggingface.co/datasets/kujiale/deep-furniture]

[4] He X, Liao L, Zhang H, et al. Neural collaborative filtering[C]//Proceedings of the 26th international conference on world wide web. 2017: 173-182.