Building a Personalized Recommendation System for E-Commerce

Introduction: This project aims to build a recommendation system for an e-commerce platform using collaborative filtering. The primary goal is to enhance user engagement by delivering personalized product recommendations based on user interactions like views, add-to-cart actions, and transactions. The model focuses on using Singular Value Decomposition (SVD) for collaborative filtering to factorize the user-item interaction matrix and make accurate predictions.

Data Collection and Preprocessing: Data sets were gotten from Kaggle, and they include:

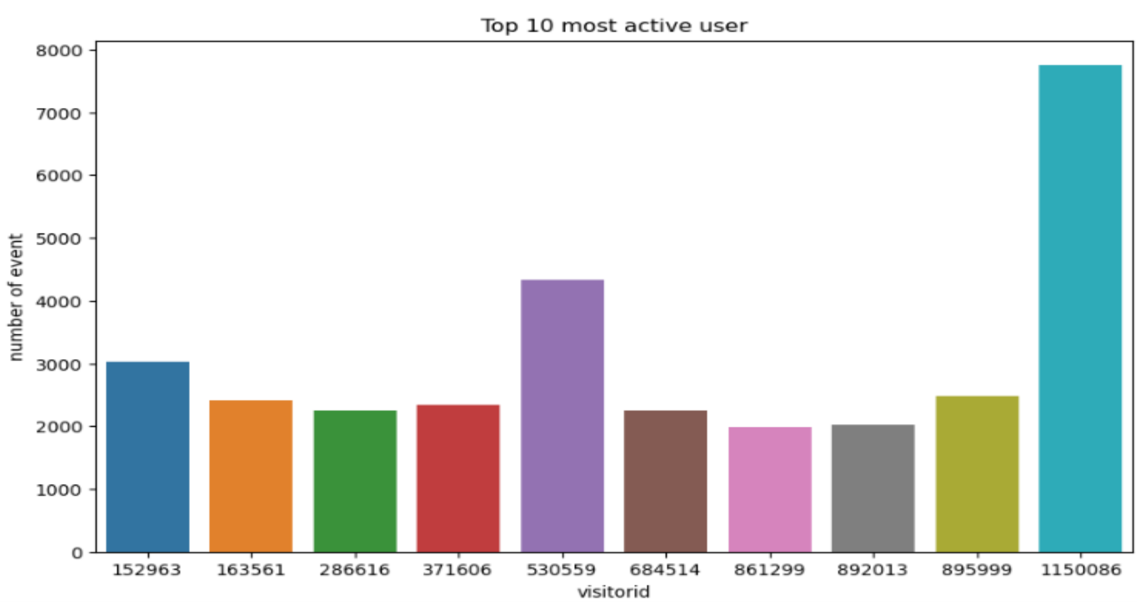
1. category_tree.csv: Contains the category hierarchy for items.

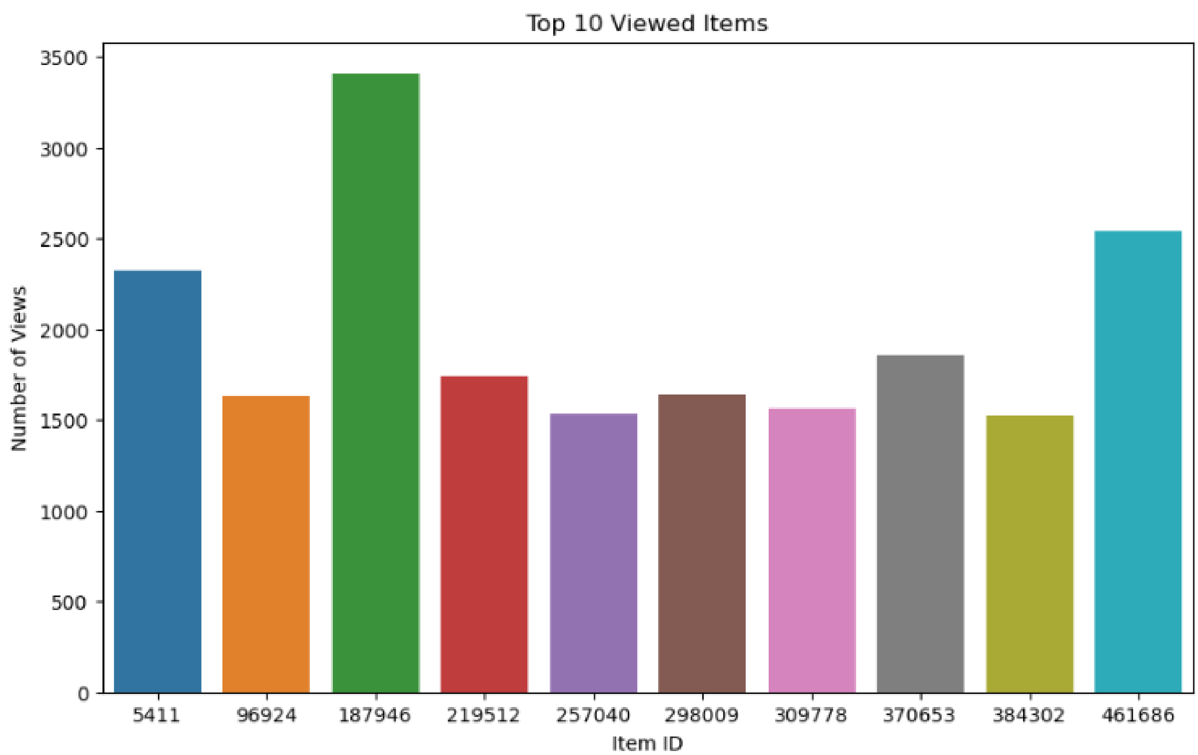
2. events.csv: Contains user interaction events (views, add to cart, transactions).

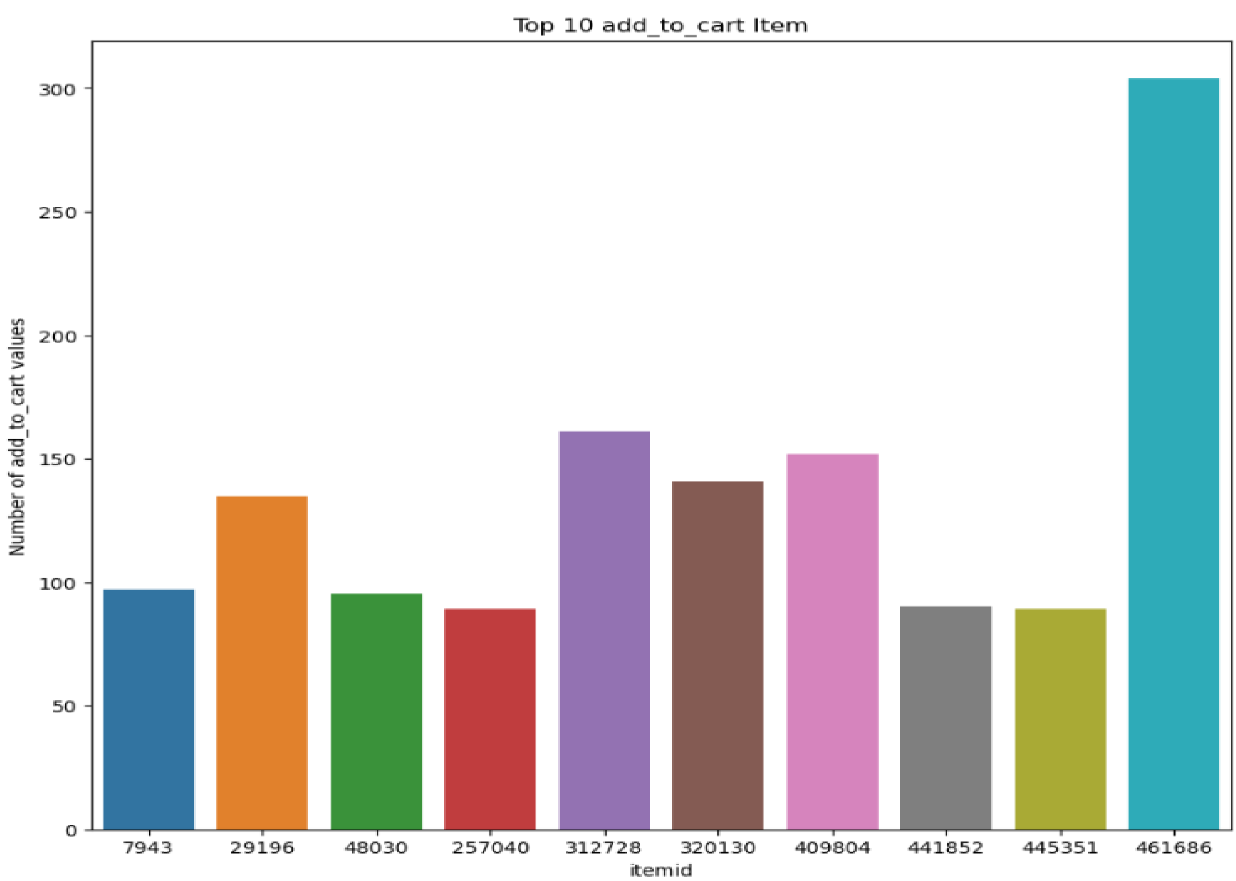
3. item_properties_part1.csv and item_properties_part2.csv: Contain metadata about item properties.

Data preprocessing steps include handling missing values, merging datasets, and converting timestamps to human-readable formats.

Exploratory Data Analysis (EDA): During EDA, we analyzed the distribution of user interactions (views, add to cart, and transactions), top items based on user engagement, and the activity patterns of the most active users. We visualized the event distribution over time and identified key trends in user behavior.




Top 10 Viewed Items
Number of Views
0
500
1000
1500
2000
2500
3000
3500
5411
96924
187946
219512
257040
298009
309778
370653
384302
461686
Item ID


Top 10 add_to_cart Item
Number of add_to_cart values
0
50
100
150
200
250
300
7943
29196
48030
257040
312728
320130
409804
441852
445351
461686
itemid

Feature Engineering: User, item, and user-item interaction features were created to enrich the dataset. These features include:

- User-based features: Item count, property count, category count, time-based features.

- Item-based features: User count, event counts (views, add to cart, transactions), conversion rates.

- Interaction-based features: Number of interactions, time between interactions.

Model Development: Used matrix factorization-based collaborative filtering using Singular Value Decomposition (SVD). The dataset was split into training (80%) and test (20%) sets, and GridSearchCV was used to tune hyperparameters like the number of factors, epochs, learning rate, and regularization term.

Hyperparameter Tuning and Best Parameters:

The best hyperparameters identified through GridSearchCV were:

- n_factors: 10

- n_epochs: 20

- lr_all: 0.002

- reg_all: 0.1

These parameters were used to train the final model.

Model Evaluation: The performance of the SVD model was evaluated using the following metrics on the test set:

- Mean Absolute Error (MAE): 0.0206

- Mean Squared Error (MSE): 0.0319

- Root Mean Squared Error (RMSE): 0.1787

The low RMSE and MAE indicate strong predictive accuracy for user-item interactions.

Recommendations: Based on these findings, the following can be done:

1. Personalized Recommendations: Use the model's predictions to provide users with personalized product recommendations, improving engagement and conversion rates.

2. Targeted Marketing Campaigns: Leverage predictive insights to offer personalized promotions and discounts, driving more effective marketing strategies.

3. Inventory Optimization: Use the model's demand predictions to optimize inventory levels, ensuring efficient stock management and reducing the risk of over- or under-stocking.

Conclusion: Collaborative filtering recommendation system using SVD was built which entails the predicting of user-item interactions with a high degree of accuracy, as shown in the low RMSE and MAE scores.

Future Work:

1. Explore hybrid recommendations systems by combining collaborative filtering with content-based methods.
2. Addressing cold start problems for new users and items.
3. Investigate real-time recommendations systems to capture dynamic user behavior.