Item-Based Recommender System For E-commerce

By: Ayodeji Ejigbo

Project Objective

Objective:

Develop an item-based recommendation system to suggest items to users based on the similarity of items they have previously interacted with.

Use Case:

Suitable for e-commerce platforms with a wide range of items where user preferences need to be predicted.

Dataset Overview

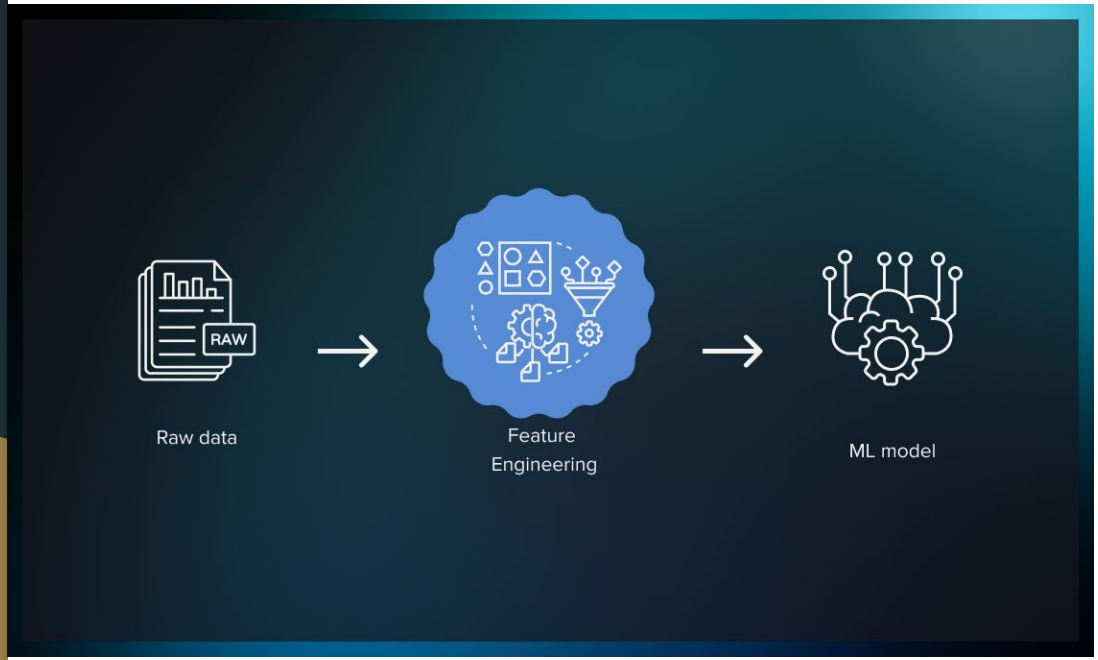
Dataset Description: The Amazon Reviews data repository contains various datasets. For this case study, we are utilizing the Electronics dataset.

- **Variables**:
  - `userid`: Unique identifier for each user.
  - `itemid`: Unique identifier for each item.
  - `rating`: Rating given by a user to an item, typically on a scale (e.g., 1 to 5).
  - `timestamp`: Time at which the rating was recorded (optional).
- **Data Source**: This is a Kaggle Dataset
- Data Shape: 1.2M rows and 100k columns.

Why Item Based Collaborative-Filtering?

- Scalability with Large User Bases
- Stability and Accuracy
- Ease of Recommendation
- Robustness to User Changes

Feature Engineering & Data Cleaning

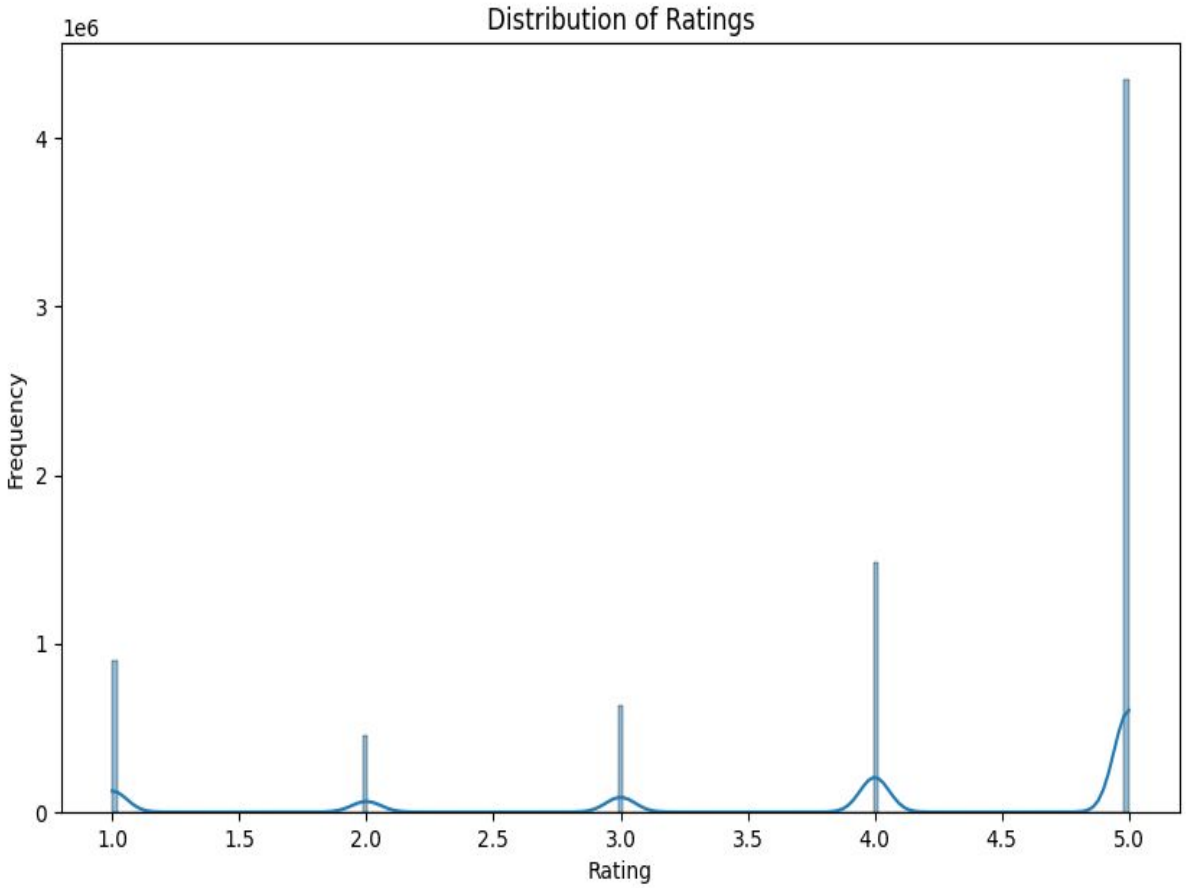


- Feature Engineering:
  - `user_avg_rating:` This shows the average rating per user.
  - `Product_avg_rating:` This shows the average rating per product.
- Data Cleaning:
  - `Missing Values`

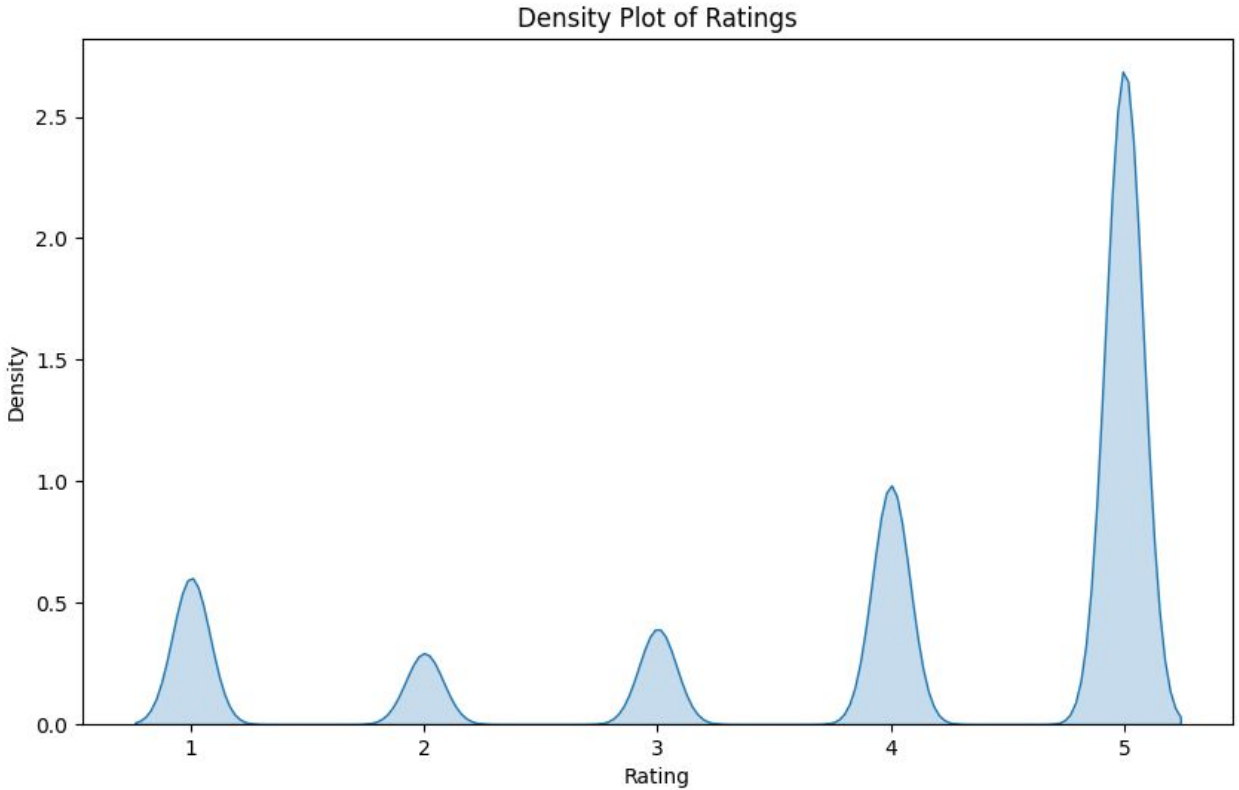
Exploratory Data Analysis

Exploratory Data Analysis

Skewness: 1.21

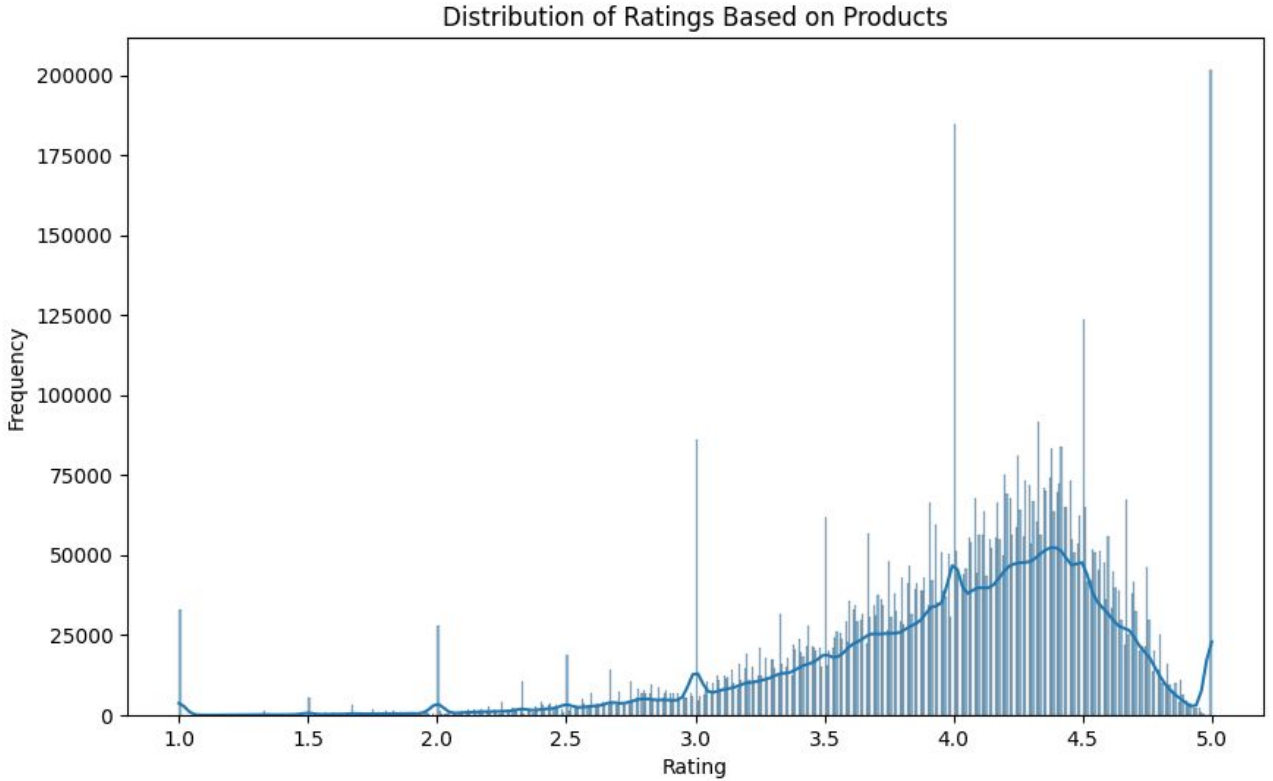


Exploratory Data Analysis



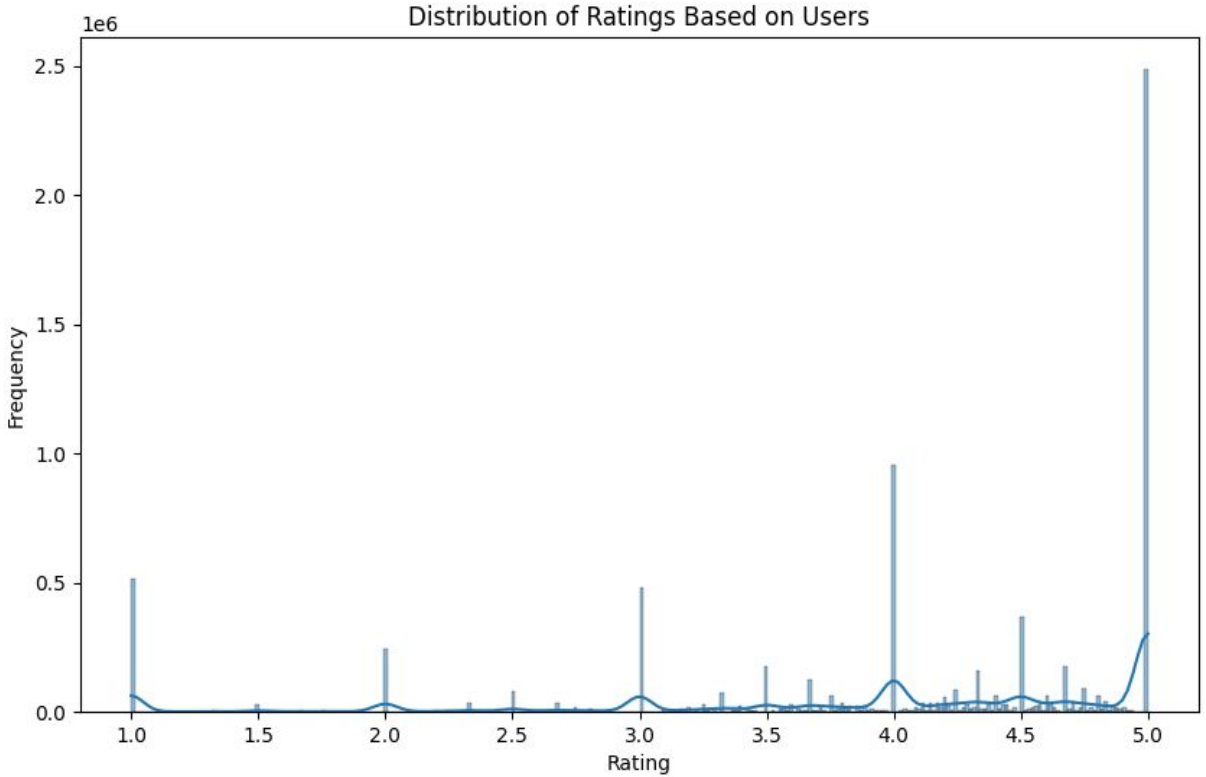
Exploratory Data Analysis

Skewness: 1.36

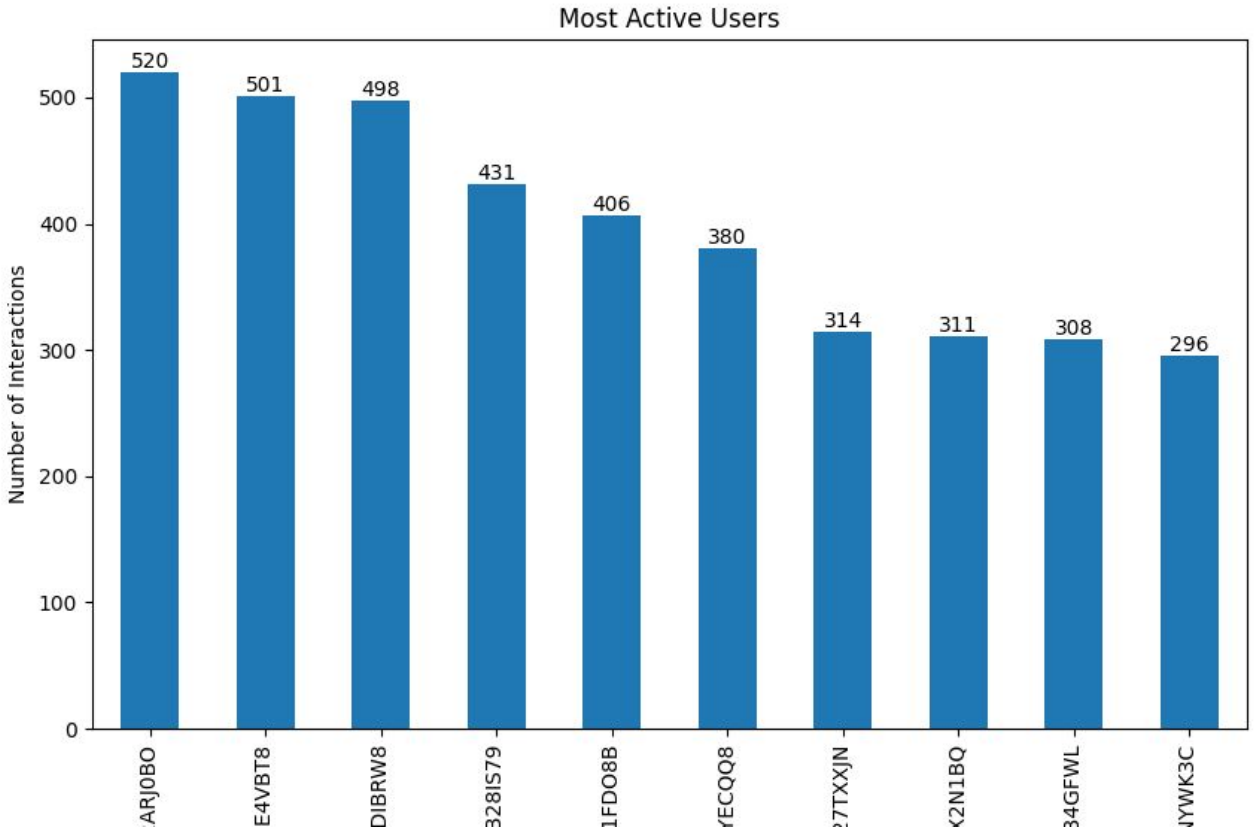


Exploratory Data Analysis

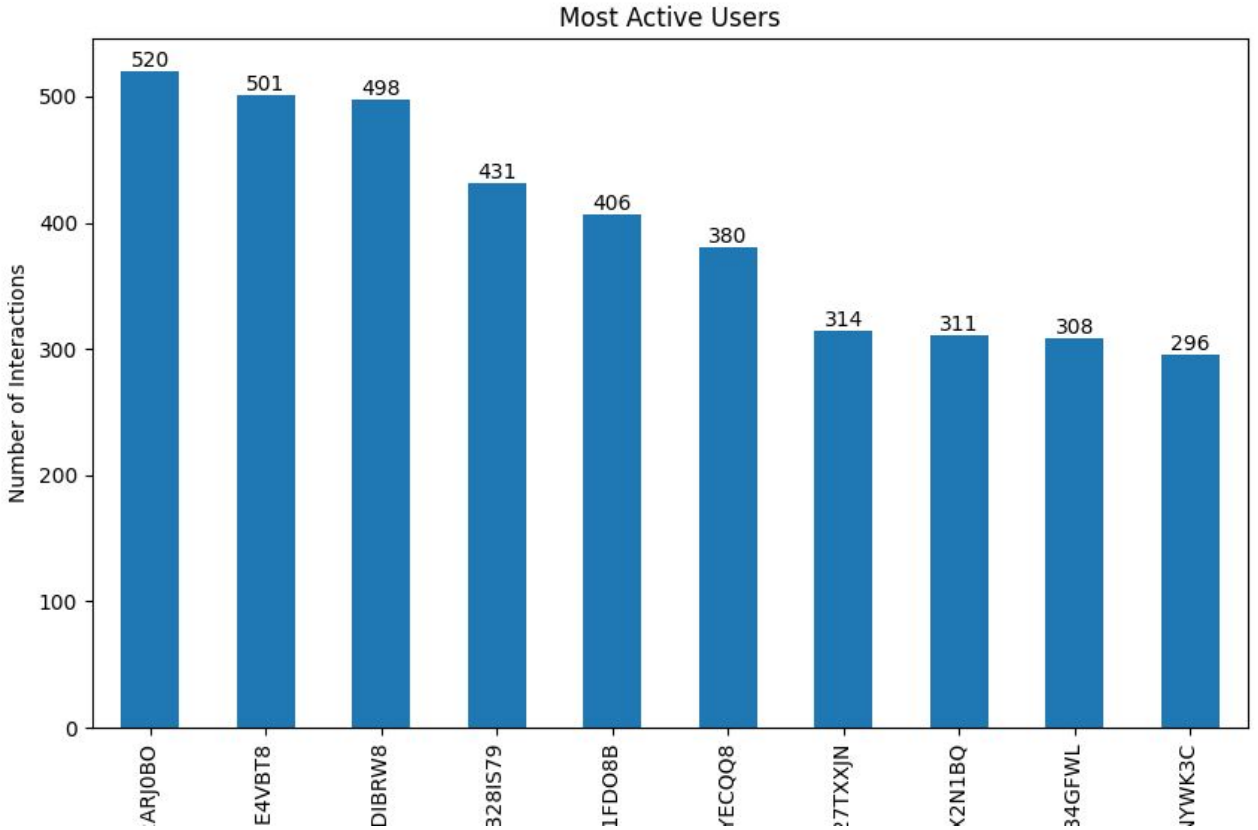
Skewness: 1.34



Exploratory Data Analysis



Exploratory Data Analysis

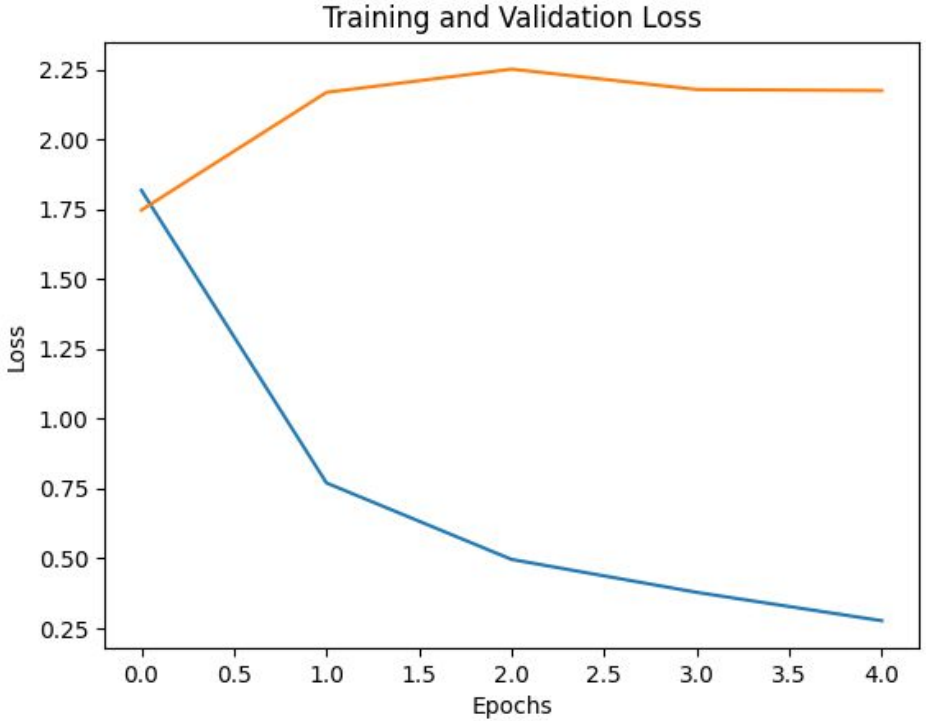


Model Development

MODELS

- **Deep Learning Models**
  - Neural Collaborative Filtering (NCF)
- **Matrix Factorization Techniques:**
  - Singular Value Decomposition (SVD)
- **K-Nearest Neighbors (KNN)**
  - KNNBasic

Neural Collaborative Filtering (NCF)



MODEL ANALYSIS

	RMSE	MAE	PRECISION	RECALL	F1-SCORE
NCF	1.47	1.033	0.77	0.79	0.78
SVD	1.30	1.031	0.77	0.93	0.84
KNNBasic	1.40	1.095	0.74	0.99	0.84

Conclusion

- **SVD**: Best overall performance with lowest RMSE and MAE; strong balance in precision and recall.
- **KNNBasic**: Excels in recall, making it ideal for scenarios prioritizing recall.
- **NCF**: Needs improvement in accuracy; higher RMSE and lower recall.

Recommendations and Future Work

Recommendations and Future Work:

- **Adopt SVD**: Use as the primary algorithm for accurate recommendations.
- **Use KNNBasic**: Apply in scenarios where recall is critical.
- **Enhance NCF**: Focus on reducing RMSE and improving recall.
- **Future Enhancements**:
  - Address cold-start with hybrid models.
  - Explore scalability through distributed computing.
  - Incorporate user feedback for continuous improvement.