# Cross Platform Recommendation System

Hyunjin Kim
6/7/2024

# Overview

Users consume a diverse array of content across multiple platforms.

Most existing recommendation systems are confined to single platforms.

We aim to bridge the gap by developing a recommendation system that integrates user interactions across multiple platforms.

# Project Formulation

Project Goals: Combine user interaction data from different platforms

Develop Predictive Model: Create machine learning models capable of providing personalized recommendations across these platforms.

Optimize Models: Fine-tune models for optimal results

# Datasets

In order to develop our model we used three distinct datasets:

Books Dataset:
Source: Kaggle
Key Feature: Book Ratings, Book Titles

Movies Dataset:
Source: MovieLens
Key Feature: Movie Ratings, Movie Title

Music Dataset:
Source: Zenodo
Key Feature: Artist Name, Number of Plays

# Creating Synthetic Users

To simulate real-world scenarios synthetic users were created by merging the three datasets.

This approach balanced the datasets and provided a comprehensive view of user preferences across all three platforms

# Key Takeaways from Data Cleaning

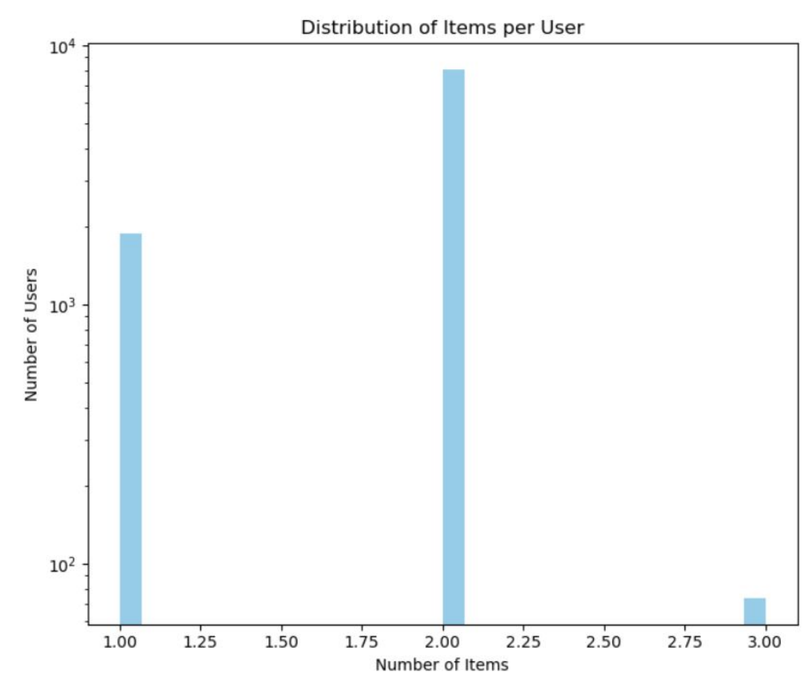
1. High Sparsity: The user item interaction matrix showed a high level of sparsity

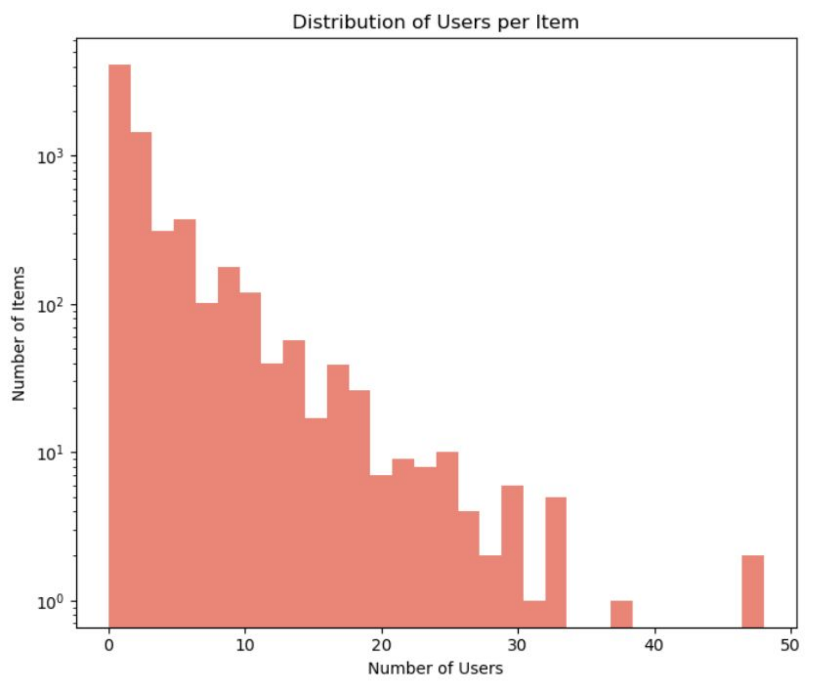
2. Data Imbalance: There was an imbalance in user and item interactions with some users and items having many interactions and some having very few to none.

3. Standardization and Consistency: To handle the different scales of interaction, all interactions were standardized to binary values. This ensured consistency across datasets and simplified the modeling.

# Exploratory Data Analysis

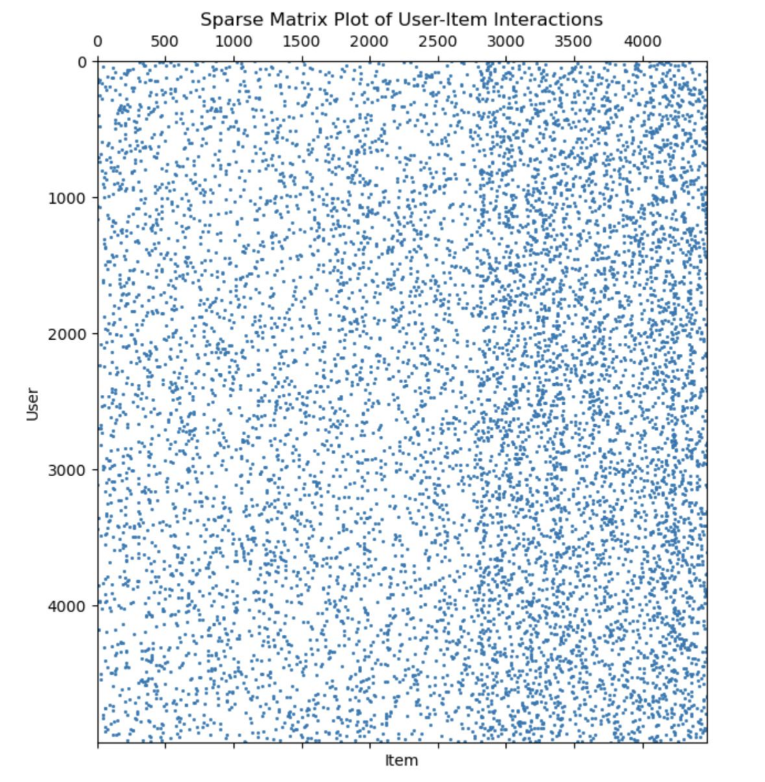
Distribution Analysis



# Exploratory Data Analysis

Visualization of user item interaction matrix




Key Takeaways:

The matrix is predominantly sparse, indicating that most users have interacted with only a small subset of items.

The scatter of interactions across the matrix shows a diverse range of user behaviors. Some users interact with a wide variety of items, while others show very limited interactions.

# Exploratory Data Analysis

## Summary Statistics:

```
Interaction Count Distribution:
0       2377
1       2832
2        677
3        337
4        219
5        140
6         75
7         55
8         38
9         25
10        29
11        12
12         9
13         8
14        11
15         3
17         3
18         3
21         1
27         1
```

```
Users after filtering: 5000
Items after filtering: 4478
Total interactions after filtering: 9123
Average interactions per user after filtering: 1.82
Average interactions per item after filtering: 2.04
```

# Model Selection

1.) Singular Value Decomposition (SVD)
Rationale: Captures underlying patterns in user preferences and item characteristics which makes it effective for sparse matrices

2.) Alternating Least Squares (ALS)
Rationale: Another matrix factorization technique. Handles implicit feedback well by assigning confidence levels to interactions.

3.) Deep Learning:
Rationale: can capture complex, non-linear relationships between users and items, and can model complex interactions and dependencies.

# Model Results

| Model | RMSE | Precision | Recall | F1-Score |
|---|---|---|---|---|
| ALS | .99 | 1 | 0 | 0 |
| SVD | .99 | 1 | 0 | 0 |
| Deep Learning | .000004 | 1 | 1 | 1 |

Key Takeaways:

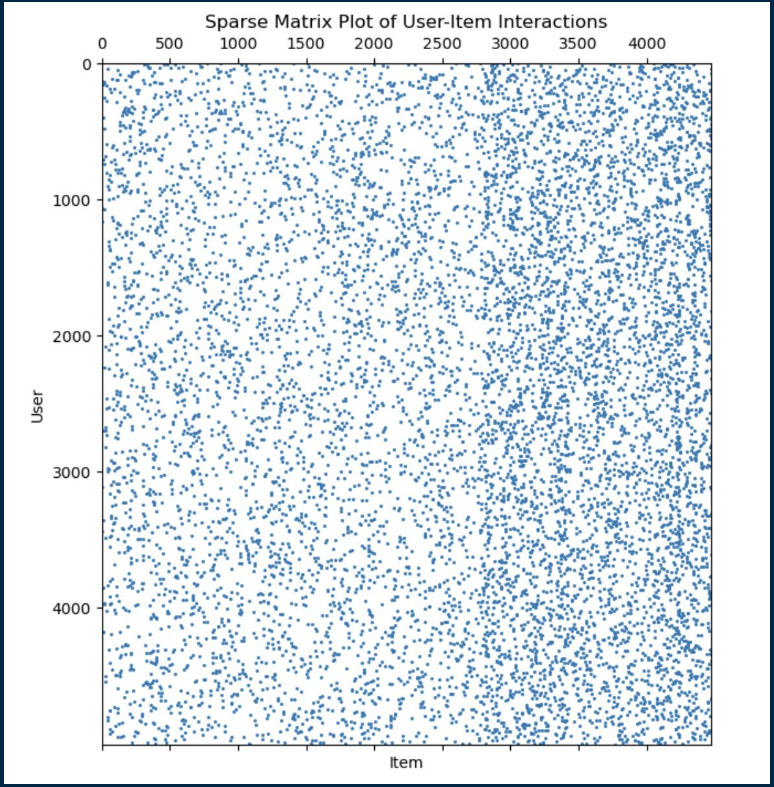
- The ALS and SVD models exhibit perfect precision but fail in recall and F1-score.
- The deep learning model overfits the training data, as indicated by its near-perfect scores across all metrics.

# Practical Considerations and Limitations

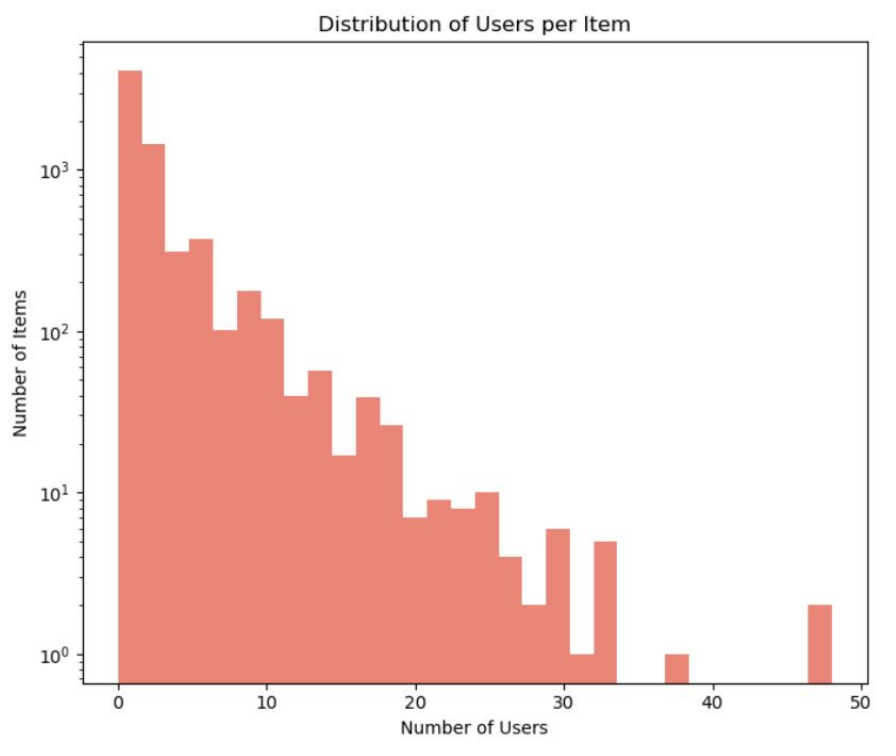
Overfitting
- As observed the models experienced overfitting due to several factors.

Data Sparsity
- The sparse user-item matrix leads to challenges in training and prediction accuracy.

Lack of content features
- The current dataset only includes user item interactions.



# Practical Considerations and Limitations




Considering there are a few items with a large number of user interactions one solution to the cold start problem could be to use these items as the initial recommendations for new users

# Suggestions for improvement and future work

- Include metadata and user demographics

```
Books_final_df.head()
```

| | User-ID | ISBN | Book-Rating | Book-Title | Book-Author | Year-Of-Publication | Publisher | Location | Age |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 99 | 0451166892 | 3 | The Pillars of the Earth | Ken Follett | 1996 | Signet Book | franktown, colorado, usa | 42.0 |

```
movies_final_df.head()
```

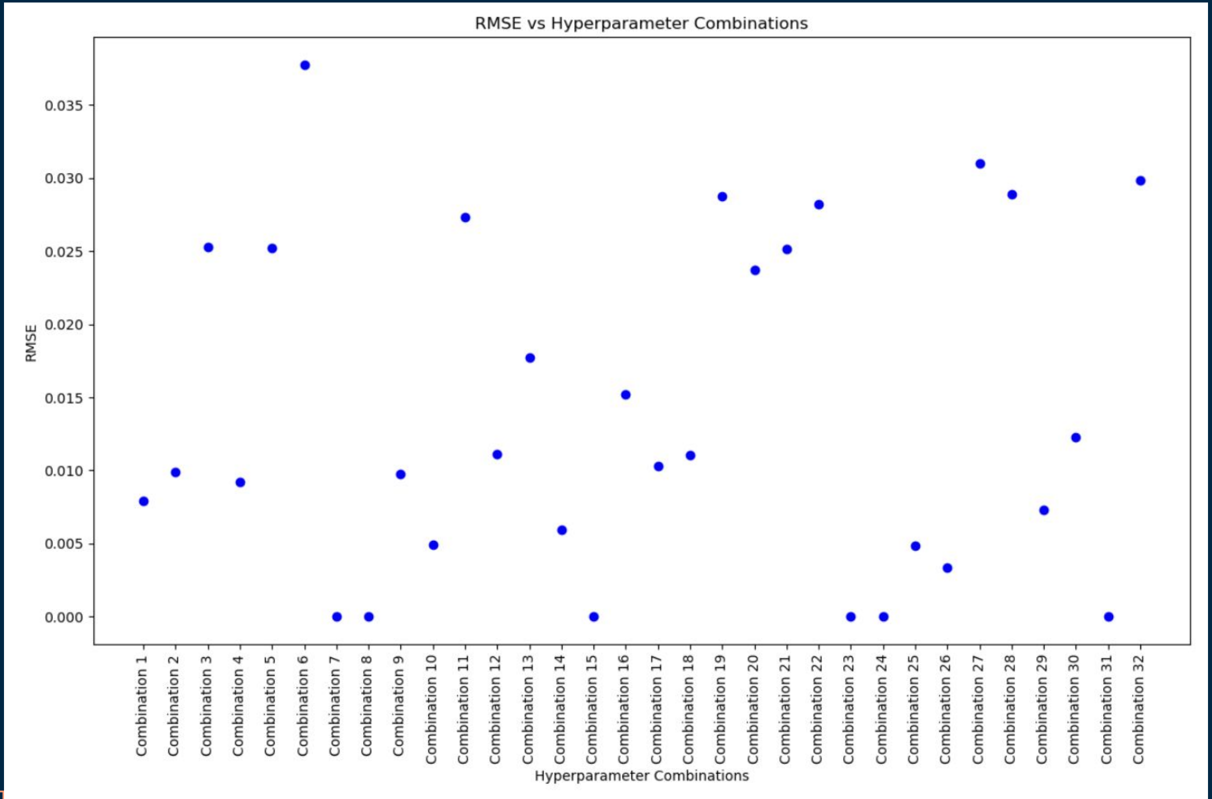
| | user_id | gender | age | occupation | zip_code | movie_id | rating | timestamp | title | genres |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | F | 1 | 10 | 48067 | 1193.0 | 5.0 | 978300760.0 | One Flew Over the Cuckoo's Nest (1975) | Drama |

```
music_final_df.head()
```

| | user_id | artist_name | plays | age | country |
|---|---|---|---|---|---|
| **0** | 00000c289a1829a808ac09c00daf10bc3c4e223b | betty blowtorch | 2137 | 22.0 | Germany |

Given that we were working with making synthetic users many of the features from our datasets were unused or scaled down.

# Suggestions for improvement and future work




The RMSE values are scattered.

Current ranges of the hyperparameters might not be appropriate.

Evaluate training and validation loss curves to identify overfitting issues.

# Conclusion and Final Thoughts

- Goal: Develop a cross-platform recommendation system for books, movies, and music.

- Methods: SVD, ALS, Deep Learning.

Key Findings:
- Deep learning model achieved the best performance after hyperparameter tuning.
- Challenges: Handling sparse data, avoiding overfitting.

# Conclusion and Final Thoughts

Next Steps and Future Work:

- Incorporate additional features for better accuracy.

- Explore ensemble methods and advanced deep learning architectures.

Final Thoughts:

- Feasibility and potential of machine learning in recommendation systems.

- Continuous improvement will refine the models.

# Thank you

- Thank you for your attention.

- Questions and feedback are welcome.