Last.FM

Unsupervised Algorithms in Machine Learning

Qiuyu Huang

2025-09

CONTENT

01

Project Background & Objectives

Why unsupervised learning for music recommendation matters, and our project goals

02

Methodology & Workflow

Overview of data preparation, models, and evaluation approach

03-08

Notebook Overview

EDA → Clustering → Recommendation → Matrix Factorization → Comparative Summary → Practical Insights & Conclusion

09

Final Takeaways

Key findings, business implications, and future directions

01 Project Background & Objectives

Background:

- Music platforms rely on **personalization** for engagement & retention
- Dataset challenge: **implicit feedback only** (play counts, no ratings)
- Key question: Can we uncover patterns and build recommendations under extreme sparsity?

Objectives:

- Explore latent patterns via **clustering**
- Evaluate **recommender baselines** under implicit feedback
- Apply **matrix factorization** to extract interpretable music themes

02 Methodology & Workflow

Data Preparation

- Clean, filter active users & items
- One-Hot encoding, scaling
- Handle imbalance, sparsity visualization

Unsupervised Models

- **Clustering:** TruncatedSVD + KMeans
- **Recommendation:** MostPopular, UserCF, ItemCF, SVD, NMF
- **Matrix Factorization:** SVD vs. NMF (RMSE comparison)

Evaluation Metrics

- **Clustering:** Silhouette Score
- **Recommendation:** Precision@K, Recall@K, Hit Rate
- **Matrix Factorization:** RMSE, interpretability

Interpretation & Insights

- From evaluation metrics → to **business-level insights**

03-08 Instructions

Sections 03–08 are shown directly in the Jupyter Notebook:

03
EDA
Sparsity, WordCloud, user/track distributions

04
Clustering
User & track clustering (Silhouette results)

05
Recommendation
Baselines & evaluation

06
Matrix Factorization
NMF vs. SVD, RMSE & interpretability

07
Comparative summary
Clustering vs. Recommendation vs. Factorization

08
Practical Insights & Conclusion
Translating findings into applications, highlighting limitations, and outlining future work.

09 Final Takeaways

Key Findings

- User clusters weak, but track clusters meaningful
- NMF outperformed SVD, revealing interpretable latent structures

Business Value

- Segment heavy vs. light listeners for retention
- Playlist generation & discovery from track-level themes

Future Directions

- Technical: ALS, BPR, Transformers
- Data: enrich with lyrics, genres, audio embeddings
- Business: recall → re-rank → interpretation loop

Unsupervised learning gave us valuable insights — weak user grouping, useful track themes, and NMF as the best balance between accuracy and interpretability.