

---

**Music Recommender System**  
Intermediate Report  
Automated Learning and Data Analysis

**Ajay Ramani**  
**Bhargava Pejakala Kakarnaya**  
**Kristy Patel**  
**Sandesh Veerapur**  
**Shini Arora**

---

---

## MUSIC RECOMMENDER SYSTEM

The Music Recommender System aims to predict which songs a user will listen to. The idea is to recommend the most popular songs to the user learning from his listening history. The aim is to suggest a set of songs to a user given half of its listening history and complete listening history of other 1 million people. This task is more properly described as a top-N recommendation task.

### DATASET

This system makes use of the Million Song Dataset (MSD) which is a freely-available collection of audio features and metadata for a million users. The data available in MSD is in the form of <user ID, song ID, play count> triplets. The full listening history is available for about a million users which serves as the training set and the validation and test sets contain half of the listening history of about 110K users.

### APPRAOCH

Collaborative Filtering (CF) is a technology which uses the items by user matrix to discover other users with similar tastes as the active user for which we want to make the prediction. The fundamental assumption of CF is that if users X and Y rate n items similarly, or have similar behaviors (e.g., buying, watching, listening), and hence will rate or act on other items similarly. The intuition is that if other users, similar to the active user, already purchased a certain item, then it is likely that the active user will like that item as well. **Memory-based collaborative filtering** is a technique to predict the item based on the entire collections of previous ratings. Every user is grouped with people with similar interests, so that a new item is produced by finding the nearest neighbor using a massive number of explicit user votes. We plan to use on memory-based collaborative filtering approaches since they are able to deal with **large datasets** in an **efficient** and **effective** way.

Memory-based CF		
Representative techniques	Main advantages	Main shortcomings
*Neighbor-based CF(item-based/user-based CF algorithms with Pearson/vector cosine correlation) *Item-based/user-based top-N recommendations	Easy implementation are dependent on human ratings	Are dependent on human ratings
	New data can be added easily and incrementally	Performance decrease when data are sparse
	Need not consider the content of the items being recommended	Cannot recommend for new users and items
	Scale well with co-rated items	Have limited scalability for large datasets

---

In memory-based CF algorithms the entire **user-item matrix** is used to generate a prediction. There can be user-based recommendation and item-based recommendation. In **User-Based Top-N Recommendation Algorithms** for top-N recommendation algorithms identify the k most similar users (nearest neighbors) to the active user using the Pearson correlation or vector-space model and make a prediction. **Item-Based Top-N Recommendation Algorithms** use the k most similar items for each item according to the similarities to make recommendation.

**Similarity Computation:** Similarity computation between items or users is a critical step in memory-based collaborative filtering algorithms. Various methods to compute similarity or weight between users or items can be,

- Correlation-Based Similarity
- Vector Cosine-Based Similarity
- Conditional probability-based similarity

We have decided to use the cosine similarity for computing similarities for both user based and item based recommendation. For item based recommendation we are considering using the conditional probability as an alternative since we are more interested in computing how likely it is that an item will be appreciated by a user when we already know that the same user likes another item.

**Prediction and Recommendation Computation:** To obtain predictions or recommendations is the most important step in a collaborative filtering system.

- Weighted Sum of Others' Ratings
- Simple Weighted Average for item-based prediction

We plan to use simple weighted sum strategy.

## **REFERENCES**

- [1]A Survey of Music Recommendation Systems and Future Perspectives by Yading Song, Simon Dixon, and Marcus Pearce  
[2]Review Article: A Survey of Collaborative Filtering Techniques by Xiaoyuan Su and Taghi M. Khoshgoftaar