



**STEVENS**  
INSTITUTE of TECHNOLOGY  
THE INNOVATION UNIVERSITY®

# MEDIA SYNERGY: A Unified Recommendation System

**AAI 695: Applied Machine Learning**

**Team 8 Final Project**

**Prof. Jiarui Li**

## Team Members

**Ashwini Ramesh**

**Benni**

*Master's in Data Science  
Dept. Mathematical Science*

Hoboken, USA

[abenni@stevens.edu](mailto:abenni@stevens.edu)

**Gahana Nagaraja**

*Master's in Data Science  
Dept. Mathematical Science  
Hoboken, USA*

[gnagaraj1@stevens.edu](mailto:gnagaraj1@stevens.edu)

**Vaishnavi Rajendra  
Dhotargavi**

*Master's in Data Science  
Dept. Mathematical Science  
Hoboken, USA*

[vdhotarg@stevens.edu](mailto:vdhotarg@stevens.edu)



# Problem Statement

- ✓ In the era of information abundance facilitated by rapid technological advancements and widespread internet access, recommendation systems have emerged as crucial tools to address information overload. In the digital age, where individuals have unparalleled access to diverse media content, ranging from music and movies to books, the challenges of recommendation systems, such as the long-tail problem and the pursuit of serendipitous content discovery, become evident.
- ✓ To tackle these challenges and enhance user experiences, our objective is to design a unified recommendation system.



# Introduction

- ✓ In the digital age, where vast libraries of music, movies, and books are at users' fingertips, the challenge lies not just in content availability but in delivering tailored and engaging recommendations.
- ✓ Recommender systems have become indispensable tools, guiding users through the expansive realms of media consumption. This prompts the exploration and development of a unified recommender system that seamlessly integrates music, movies, and books, aiming to provide users with a comprehensive and personalized content discovery experience.
- ✓ The essence of this unified approach lies in acknowledging the diverse interests and preferences of users. By blending both popularity and similarity scores, the recommender system seeks to strike a harmonious balance.
- ✓ On one hand, it leverages the collective wisdom of the crowd, recommending popular and trending content that resonates with a broad audience. On the other, it delves into the intricacies of individual tastes, presenting users with suggestions based on the similarities observed in their past interactions.



# Related Work

- 1) Madhuri, Alekhya, MohanaVyshnavi, Aparna, Swetha and Mounika in their paper [2] have discussed and built a recommendation system using Singular Value Decomposition model which is one of the matrix factorization techniques. The quick sort algorithm is used to sort the dataset based on the keywords provided by the users after registering. User profile as well as item profile is maintained to find the “User Behaviour” which is very effective in finding the desired output.
- 2) A paper published by members of the Yildiz Technical University, Istanbul, aims to classify and recommend songs using acoustic features, extracted by digital signal processing methods and convolutional neural networks. They have conducted the study in mainly two steps: determining how features that will be used in recommendation are obtained and developing a service that recommends songs to user requests.



# K-Means Clustering

## Definition:

Unsupervised algorithm for partitioning datasets into  $K$  distinct clusters, widely used in data segmentation, customer profiling, image compression, and anomaly detection.

## Methodology:

1. Initialization: Randomly select  $K$  centroids.
2. Assignment: Assign each point to the nearest centroid.
3. Update Centroids: Recalculate centroids based on mean of assigned points.
4. Repeat: Iterate until centroids converge or a set number of iterations.

## Limitations:

Efficient but sensitive to initial centroids, outliers, and requires pre-specifying  $K$ .



# K-Means Clustering Implementation

Steps Involved in Implementation of K-Means Clustering:

1. Data Pre-processing
2. Dataset Overview
3. Feature Correlation Analysis
4. Exploratory Visualization
5. Clustering Analysis
6. Dimensionality Reduction and Visualization
7. Personalized Song Recommendation

# K-Means Clustering Outcome

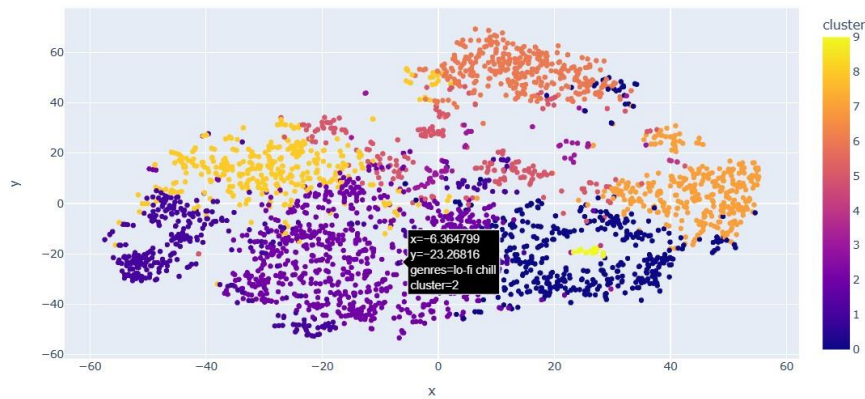


Fig (i) : Displays the genre clustering

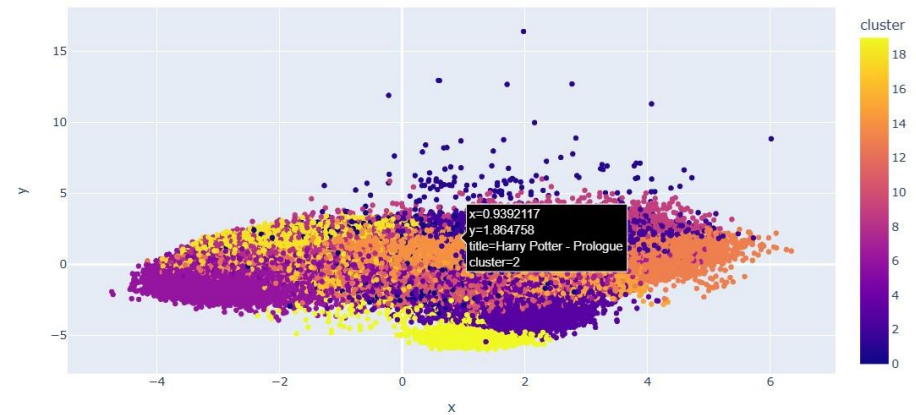


Fig (ii) : Displays the song clustering

Based on the analysis, I recommend you listen to 'Happy' from Cluster 0. Enjoy!

Fig (iii) : Figure showing the song recommendation



# Collaborative filtering based approach

Collaborative filtering is a popular approach in recommender systems that makes automatic predictions (filtering) about the preferences of a user by collecting preferences from many users (collaborating). The underlying idea is to recommend items to a user based on the preferences and behaviors of other users who have similar tastes or preferences. Collaborative filtering can be broadly classified into two main types: user-based collaborative filtering and item-based collaborative filtering.

- ✓ Calculate the similarity between items. This is often done using metrics like cosine similarity or Pearson correlation coefficient.
- ✓ Determine which items are most similar to the ones the user has interacted with based on user-item interactions.
- ✓ Predict the ratings for items that the target user has not yet interacted with by combining the ratings of similar items.
- ✓ Weigh the ratings of similar items based on their degree of similarity to the items the user has already interacted with.





## Data Preprocessing:

### Cosine Similarity

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space based on the cosine of the angle between them, resulting in a value between -1 and 1. The value -1 means that the vectors are opposite, 0 represents orthogonal vectors, and value 1 signifies similar vectors.

```
In [66]: popularity_df
```

Out[66]:

	original_title	authors	rating_x	rating_y
0	Still Life with Woodpecker	Tom Robbins	99	4.777778
1	The Beautiful and Damned	F. Scott Fitzgerald	100	4.660000
2	Villa Incognito	Tom Robbins	97	4.618557
3	Peter and the Shadow Thieves	Dave Barry, Ridley Pearson, Greg Call	70	4.557143
4	The Taste of Home Cookbook	Janet Briggs, Beth Wittlinger	100	4.550000
5	A People's History of the United States: 1492 ...	Howard Zinn	100	4.540000
6	Hard Times: For These Times	Charles Dickens	99	4.535354
7	Girl with a Pearl Earring	Tracy Chevalier	100	4.530000
8	Deception Point	Dan Brown	100	4.500000
9	El amor en los tiempos del cólera	Gabriel García Márquez, Edith Grossman	94	4.500000
10	Amsterdam	Ian McEwan	95	4.494737

Fig Shows the top 50 book recommendations based on popularity



---

Top 5 book recommendations based on the similarity is:

The Secret Garden

Wild Swans: Three Daughters of China

Message in a Bottle

The Clan of the Cave Bear

The Time Machine

Fig Shows the collaborative filtering-based technique used in book recommendation system.



# Linear Regression model

Linear regression is a statistical method used to model the relationship between a dependent variable (what you want to predict) and one or more independent variables (the features or factors that may influence the prediction) by fitting a linear equation to observed data.

$$Y = b_0 + b_1X^1 + b_2X^2 + \dots + b_nX^n$$

where:

- $Y$  is the dependent variable,
- $X^1, X^2, \dots, X^n$  are the independent variables,
- $b_0$  is the y-intercept,

# Linear Regression Model Implementation



- ✓ The dataset was loaded into the dataframe using pandas .
- ✓ The dataset is preprocessed, NaN and infinite values are removed, and selected numerical features undergo a log transformation.
- ✓ Targets and Covariates are been initialized, filtering the movies based on the input.
- ✓ The features are extracted and the models are instantiated.
- ✓ The models are trained and the recommendation are displayed.



# Linear Regression Model Output

Movie Recommended by Highest vote\_count on various Algorithm models:

Recommendation based on Linear Regression:

	original_title	vote_count	Linear_Regression_Prediction
65	The Dark Knight	4.07929	9.000849

Recommendation based on Logistic Regression:

	original_title	vote_count	Logistic_Regression_Prediction
65	The Dark Knight	4.07929	1

Recommendation based on Random Forest Regression:

	original_title	vote_count	Random_Forest_Regression_Prediction
65	The Dark Knight	4.07929	8.868177

Recommendation based on Random Forest Classification:

	original_title	vote_count	Random_Forest_Classification_Prediction
65	The Dark Knight	4.07929	1

Performance Metrics:

Linear Regression MSE: 0.2942413330791596, R-squared: 0.5888442367519282

Logistic Regression MSE: 0.14763014763014762, R-squared: -0.0024350649350650677

Random Forest Regression MSE: 0.20408395404788454, R-squared: 0.7148249261409239

Random Forest Classification MSE: 0.13053613053613053, R-squared: 0.11363636363636354

Fig Shows the movie recommendation based on the highest count,  
by the genre specified by user



# Conclusion

- ✓ In conclusion, the development of a unified recommender system encompassing music, movies, and books introduces a dynamic and versatile platform that aims to enhance user engagement and satisfaction.
- ✓ By incorporating both popularity and similarity scores, the system delivers a nuanced approach to recommendations, striking a balance between widely appreciated content and personalized user preferences.
- ✓ This approach ensures that users not only encounter popular and trending items but also discover new and relevant content aligned with their individual tastes.