# Collaborative Filtering-based Recommendation System for Movies

Anurathi Bala, Aradhana Jayaprakash
bala.an@northeastern.edu, jayaprakash.a@northeastern.edu

## Problem Definition

Accurately build movie recommendation system that suggests movie according to user's interests by clustering based collaborative-filtering using MovieLens dataset.

## Existing Methods

- Clustering-based collaborative filtering approach with unrated movie rating by user as 0 → misinterpreted as user disliking the movie
- Assigning majority class to user based on cluster → recommend highly rated movies in that class

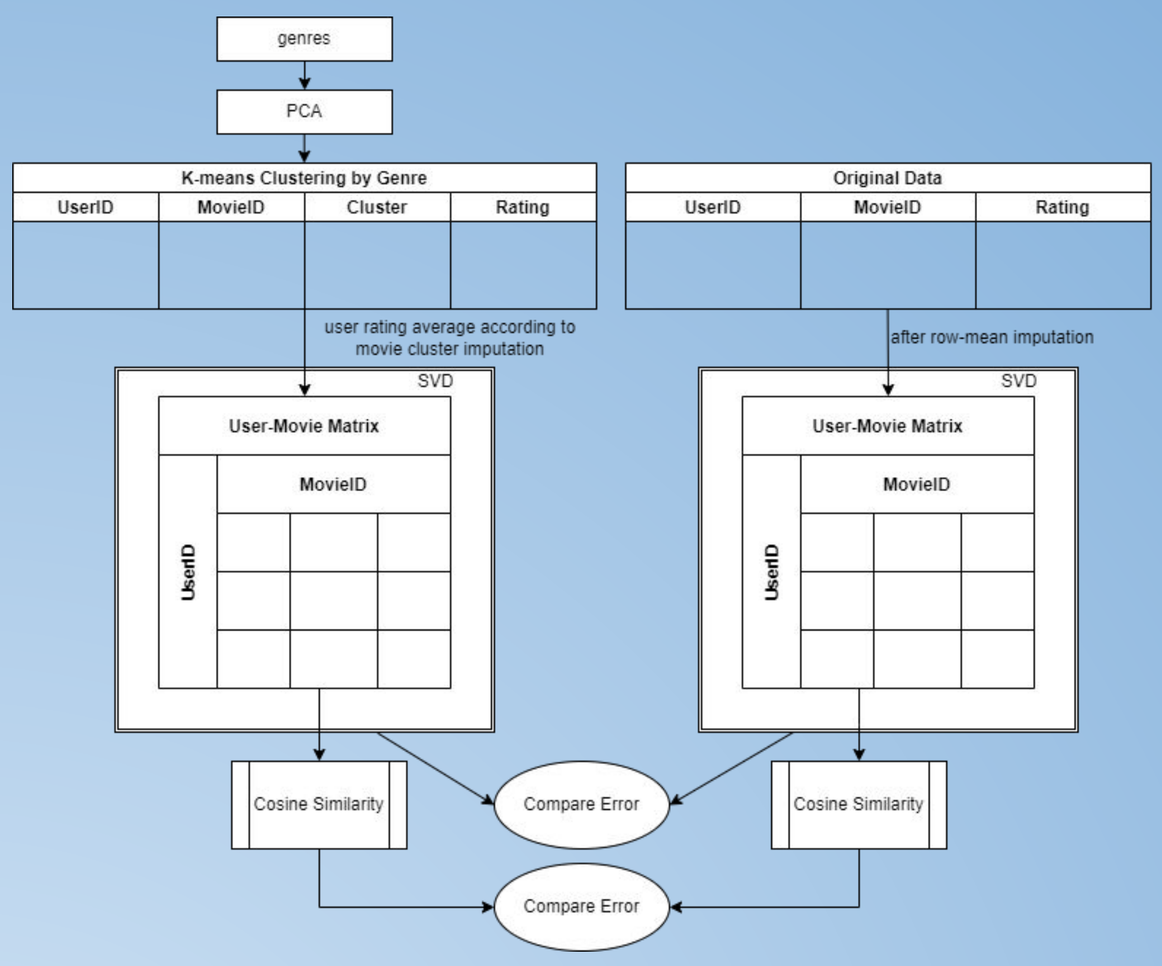
## Proposed Method

1) Data pre-processing
2) Implementing genre-based movie clustering
3) Imputing missing values of user-movie matrix with user cluster averages
4) Building Model-Based & Memory-Based Collaborative filtering model
5) Evaluating error metrics using baseline model

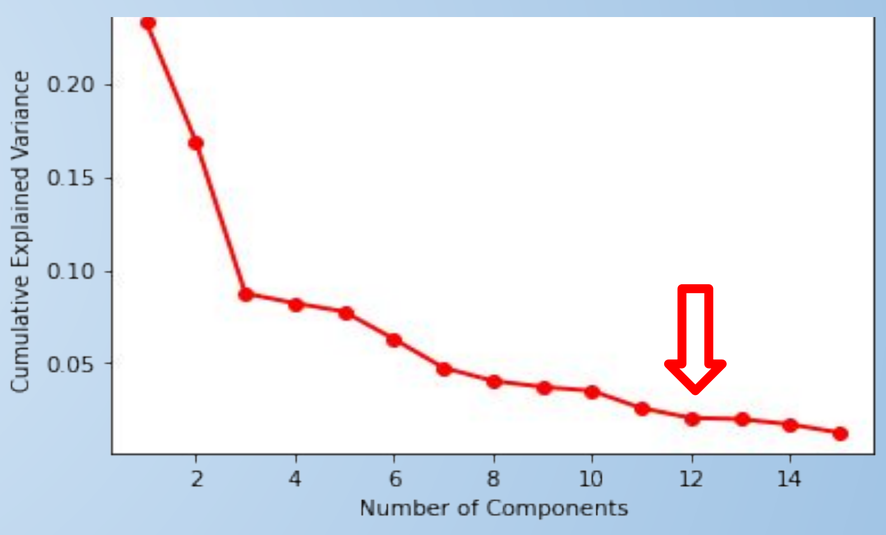
## Data Description & Experimental Setup

- MovieLens data set:
  - 2 data sets: Ratings, Movies
    - Ratings: 'UserID', 'MovieID' and 'Rating'
      - Rating: 1-5
    - Movies: 'MovieID', 'Title', and 'Genres'
  - 1,000,209 movie ratings
  - 6040 users, 3706 movies, 18 genres

## ARCHITECTURE DIAGRAM




## ELBOW METHOD




## Results

| Error Metrics | | |
|---|---|---|
| | RMSE | MAE |
| Memory-Based | 0.837 | 0.424 |
| Model-Based | 0.693 | 0.427 |

| UserID 1 Predicted Ratings Comparison | |
|---|---|
| | MovieID 1, 2, 3 Predicted Ratings |
| Model-Based with Clustering | 4.30, 4.23, 3.99 |
| Model-Based | 4.24, 4.18, 4.18 |
| Memory-Based with Clustering | 3.56, 3.54, 2.92 |
| Memory-Based | 4.19, 4.19, 4.19 |

## Discussion of Results

- Error values low for both methods → proposed approach works
- Memory-Based approach yielded lower error than the Model-Based approach by 0.837 - 0.693 = 0.144
- Model-Based approach usually outperforms that of Memory-Based → not the case here
- Clustering results show more specific rating predictions than non-clustered results

## Takeaway Points & Future Work

- Memory-Based Collaborative filtering is more accurate
- Error metrics yielded small values → proposed hybrid recommendation system approach works
- Future Work: Better missing value imputation, clustering by different features, validation of performance using active user