

# ***MOVIE RECOMMENDER SYSTEM***



**METIS**

Leo Siu-Yin

# Why A Movie Recommender System?



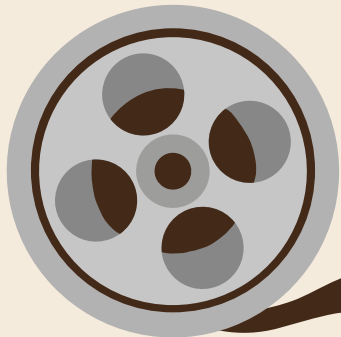
Photo by [freestocks](#) on [Unsplash](#)

A movie recommender system can be used as a playlist generator, for video or music services like Netflix or YouTube.

# Project Stages

## Preprocessing

- Load the data
- Apply Preprocessing



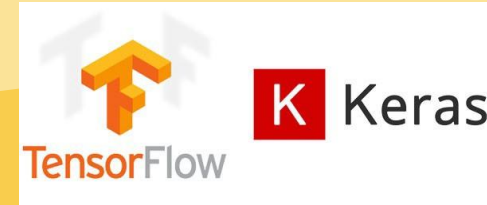
## Build Model

- Prepare train and validation data
- Create the model
- Train the model based on the split

## Results

- Plot training and validation loss
- Show top 10 movie recommendations to user

# Tools Used



# Data Set

MovieLens 100K dataset collected by GroupLens Research

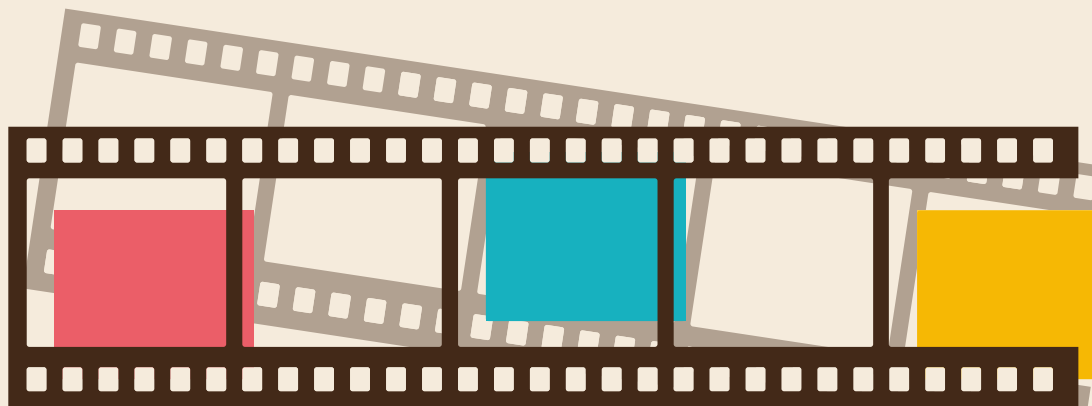
**943** users

**Each user** rated at least **20 movies**.

**100,000** ratings

**1682** movies

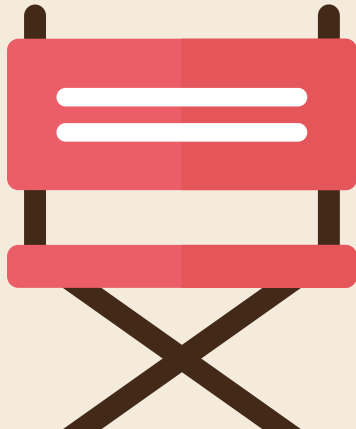
A scale of **1 to 5**  
(explicit responses)



# Models Used

## KNN

Memory Based



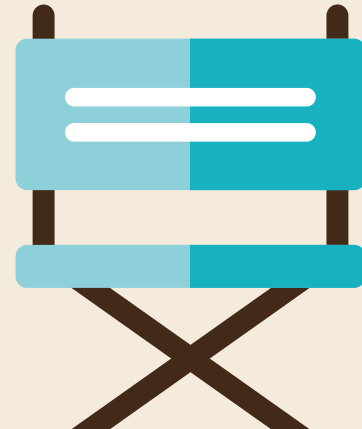
## SVD

Model Based

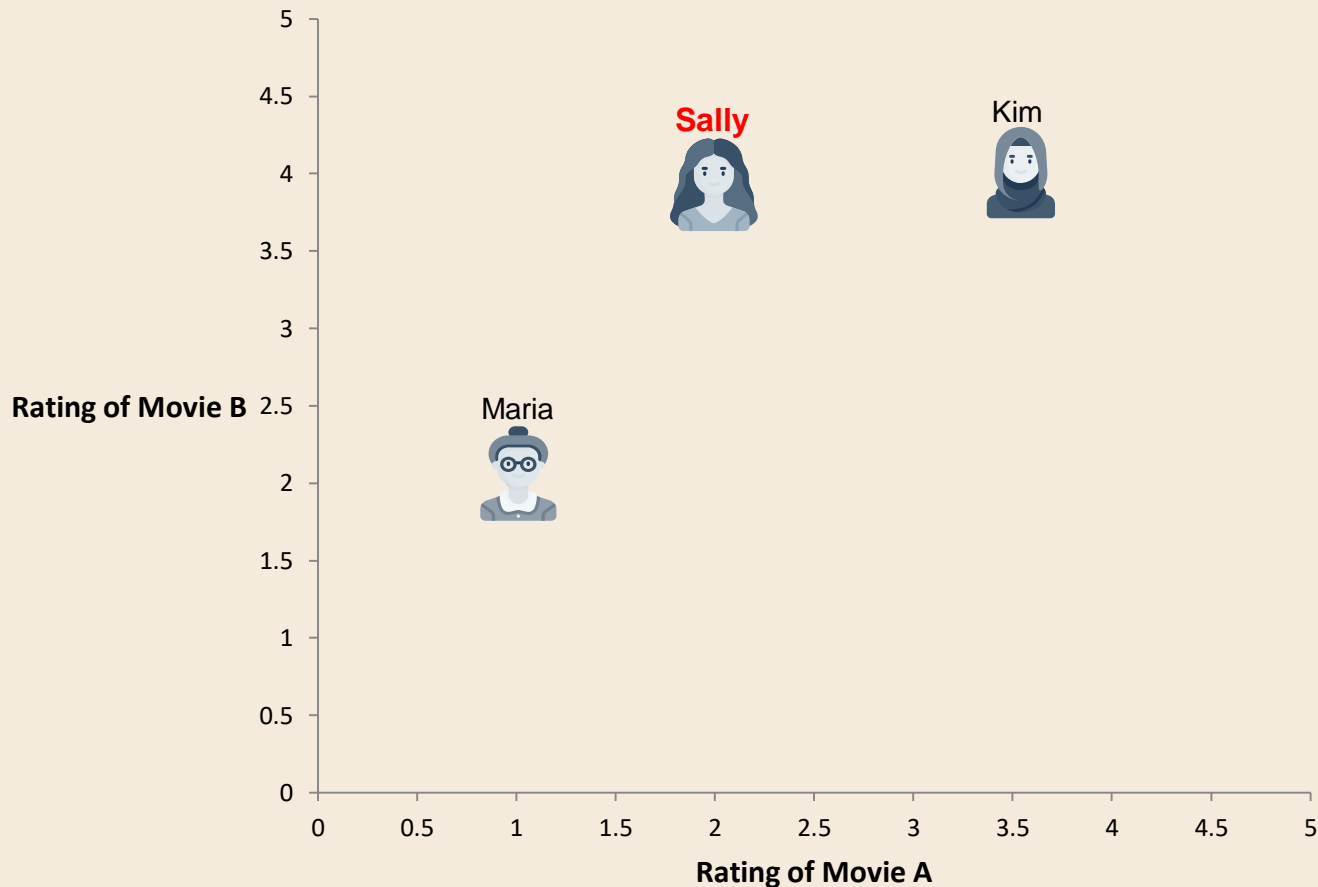


## Neural Network

Model Based



# Collaborative Based Filtering (User Based)

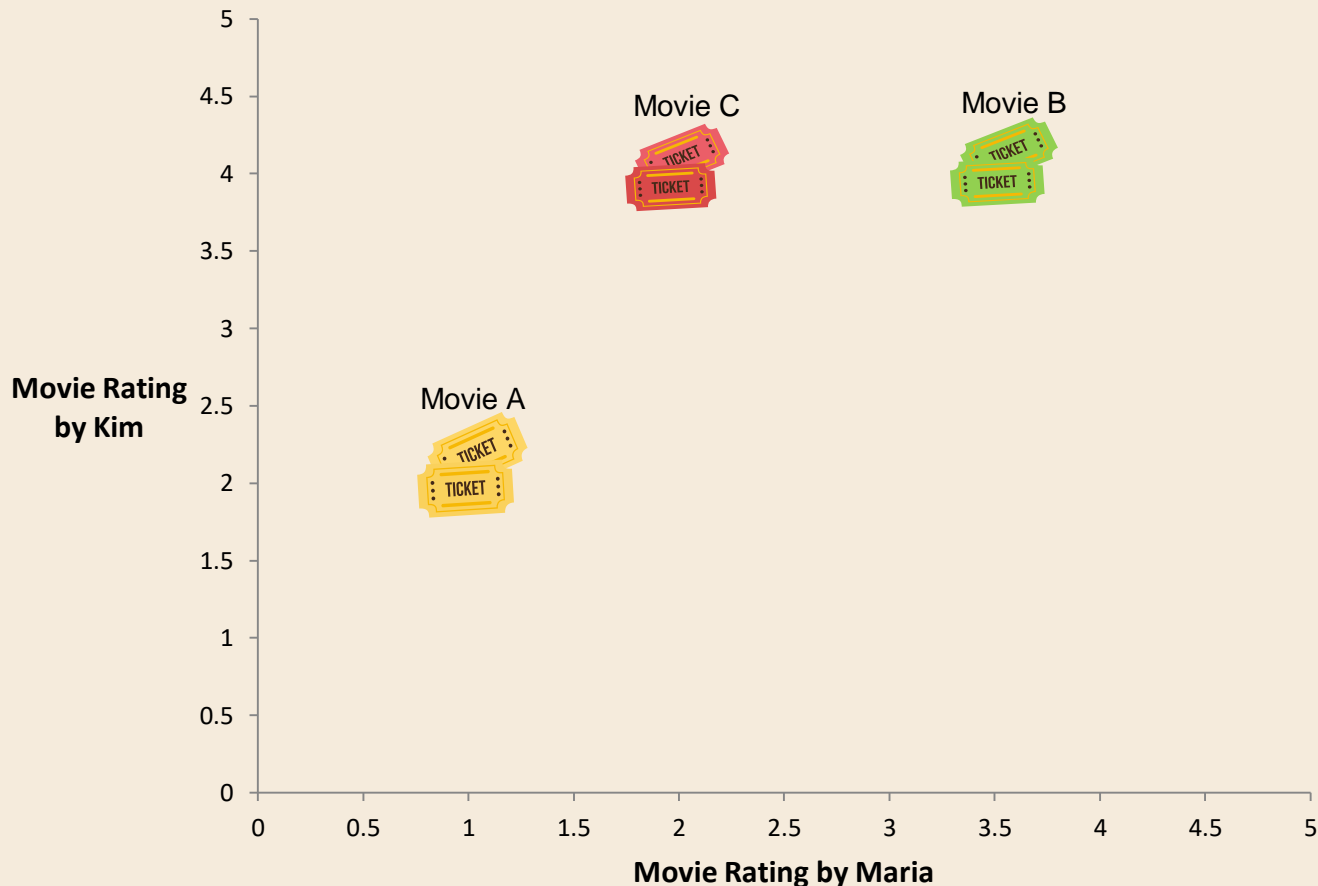


## Cosine Similarity:

Maria is a more similar user to Sally than Kim is to Sally.

Hence Maria's rating for Movie C will be used to predict Sally's rating for Movie C (which is unrated yet).

# Collaborative Based Filtering (Item Based)



## Cosine Similarity:

Based on users' ratings, Movie A is more similar to Movie C than Movie B is to Movie C.

Hence Sally's rating of Movie A will be used to predict Movie C (which is unrated by Sally yet).



# Memory-Based Collaborative Filtering- k-NN

## Model Building

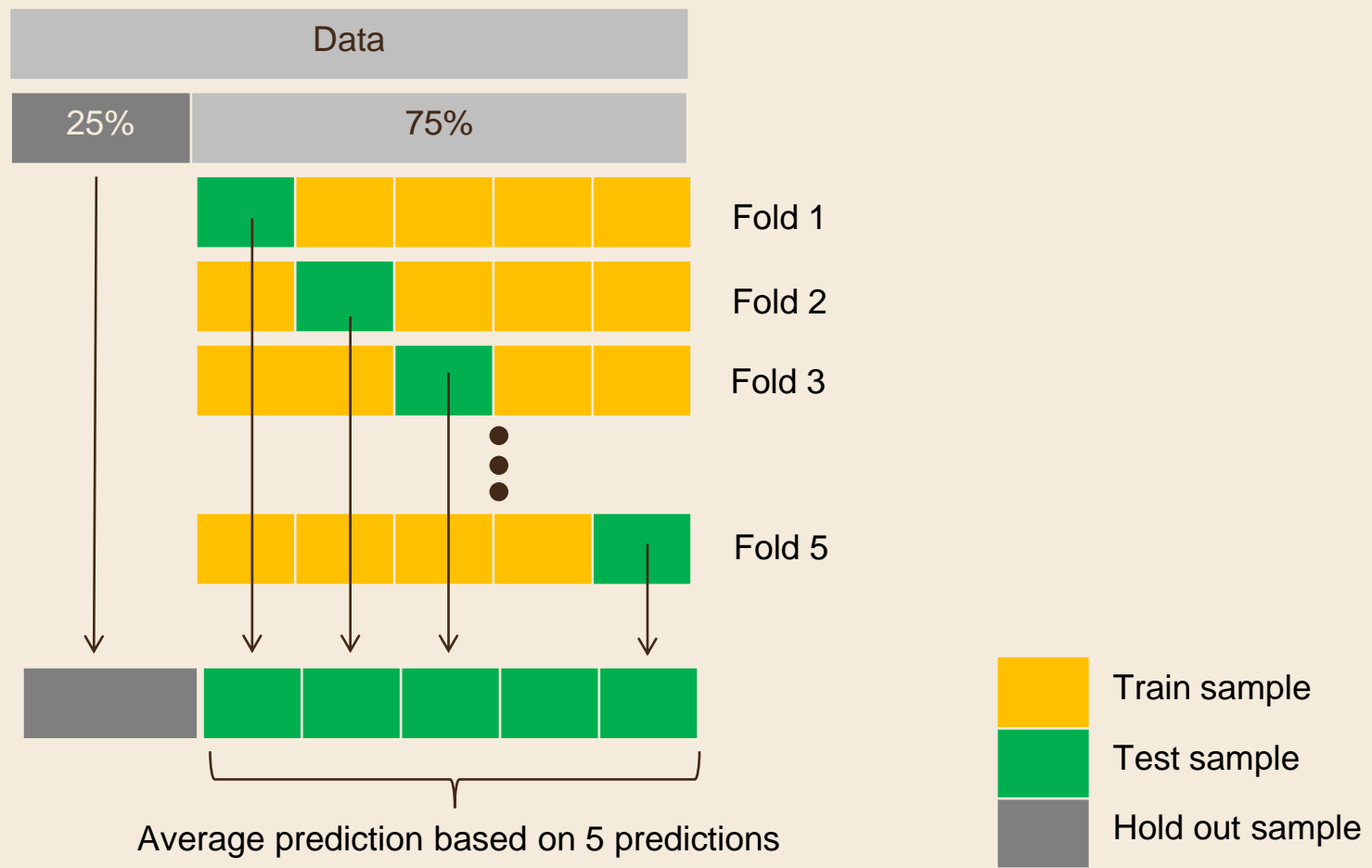
- Data is split into a 75% train-test sample and 25% holdout sample.
- Cosine similarity is chosen as the similarity measure. This computes the cosine similarity between all pairs of users (or items).
- GridSearchCV carried out over 5 -fold, is used to find the best set of similarity measure configuration (sim\_options) for the prediction algorithm, based on accuracy metrics chosen. It will find out whether user-based or item-based gives the best accuracy results
- KNNWithMeans takes into account the mean ratings of each user. Individual user preferences is accounted for by removing their biases through this algorithm.

# Model-Based Collaborative Filtering- SVD

## Model Building

- Compresses user-item matrix into a low-dimensional representation in terms of latent factors.
- Latent factors provide hidden characteristics about users and items.
- A user's interaction with an item is modelled as the product of their latent vectors.
- GridSearchCV is used to find the best configuration of the number of iterations of the stochastic gradient descent procedure, the learning rate and the regularization term.

# GridSearchCV



# Neural Network-Based Collaborative Filtering

## Data Preprocessing

- The data file that consists of users, movies, ratings and timestamp is read into a pandas dataframe for data preprocessing.
- Movies and users need to be enumerated to be used for modeling.
- Variables with the total number of unique users and movies in the data are created, and then mapped back to the movie id and user id.
- The minimum and maximum ratings present in the data are found.
- Ratings are then normalized for ease of training the model.

# Neural Network-Based Collaborative Filtering

## Model Building

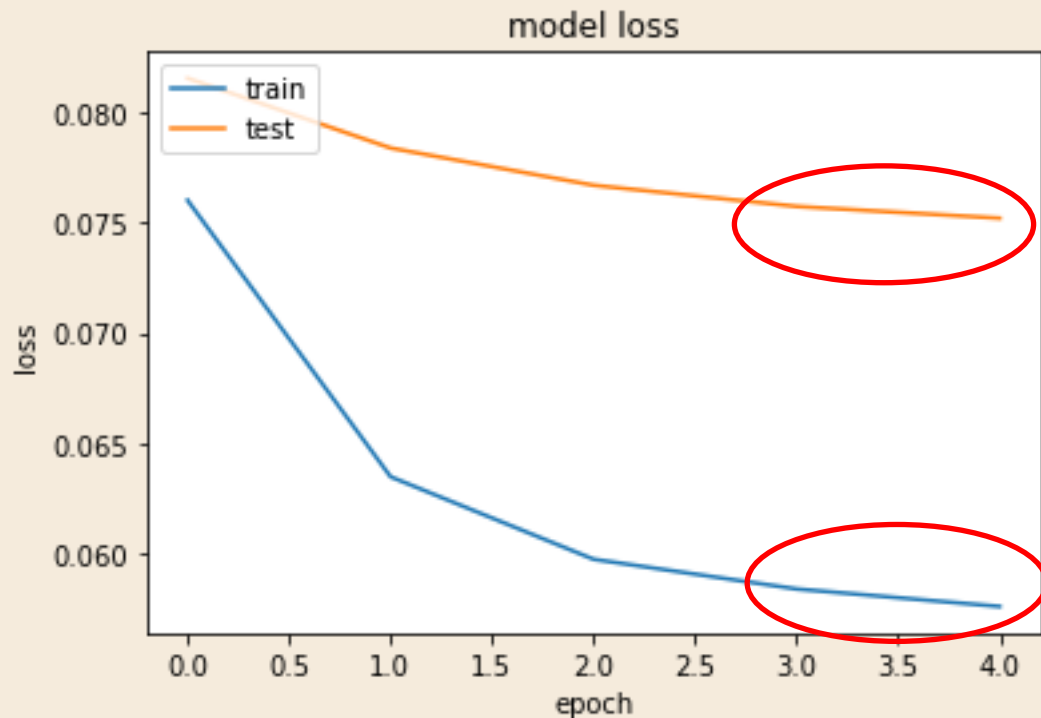
- Embeddings are used to represent each user and each movie in the data.
- These embeddings will be of vectors size  $n$  that are fit by the model to capture the interaction of each user/movie.
- Both the users and movies are embedded into 50-dimensional ( $n = 50$ ) array vectors for use in the training and test data. Training is carried out on 75% of the data and testing on 25% of the data.
- To capture the user-movie interaction, the dot product between the user vector and the movie vector is computed to get a predicted rating.
- The Adam optimizer is used to minimize the accuracy losses between the predicted values and the actual test values.

# Results

Accuracy Metric	k-NN-based	SVD-based	Neural-based
RMSE (GridSearch CV)	0.9551	0.9528	N.A.
RMSE (Holdout)	0.9402	0.9430	N.A.
MSE	0.8840	0.8889	0.0749
MAE	0.7419	0.7535	0.2242



# Neural Network – Training and Validation Loss (MAE)



## A Good Fit!

1. The plot of training loss decreases to a point of stability.
2. The plot of validation loss decreases to a point of stability and has a small gap with the training loss.

# Neural Network- Results

Showing recommendations for user: 838

@~~~~@~~~~@~~~~@~~~~@~~~~@~~~~@~~~~@

Movies with high ratings from user

@~~~~@~~~~@~~~~@~~~~@~~~~@~~~~@~~~~@

Toy Story (1995)

Godfather, The (1972)

Sound of Music, The (1965)

Empire Strikes Back, The (1980)

Princess Bride, The (1987)

Clockwork Orange, A (1971)

Evita (1996)

Titanic (1997)

Schindler's List (1993)

Mary Poppins (1964)

@~~~~@~~~~@~~~~@~~~~@~~~~@~~~~@~~~~@

Top 10 movie recommendations

@~~~~@~~~~@~~~~@~~~~@~~~~@~~~~@~~~~@

Shawshank Redemption, The (1994)

Blade Runner (1982)

Silence of the Lambs, The (1991)

Cinema Paradiso (1988)

12 Angry Men (1957)

Psycho (1960)

Graduate, The (1967)

Bridge on the River Kwai, The (1957)

Casablanca (1942)

Rear Window (1954)

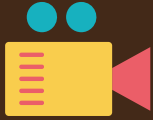


# Conclusion

Neural network based model gave the best modeling result compared to KNN and SVD models.



# Future Projects



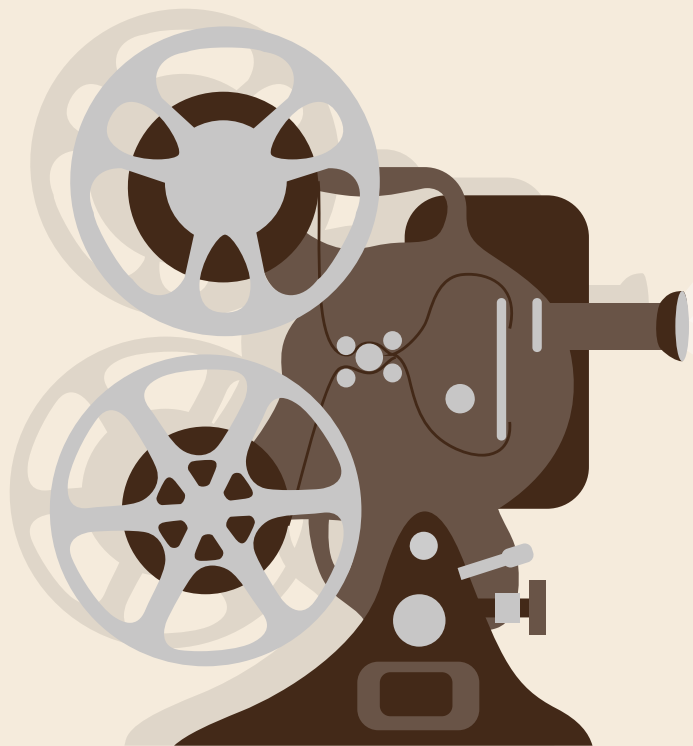
**Content Based  
Filtering**



**Hybrid memory  
based filtering**



**Explore on  
trying different  
neural network  
models.**



# THANKS

Do you have any questions?

syleo22@gmail.com

<https://www.linkedin.com/in/syleo/>

[https://github.com/syleo22/SiuYin\\_Projects](https://github.com/syleo22/SiuYin_Projects)



CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by [Freepik](#).

Please keep this slide for attribution.