

Experiment 7

A/B Testing

Instruction: Create a new .Rmd file and write code for each section in separate R code block

A/B testing compares two versions (A and B) of a system or feature to see which performs better on a key metric, like clicks or conversions, by randomly splitting users and analyzing outcome differences statistically.

Dataset: The MovieLens dataset contains four main columns. The **user** column gives each user a unique ID number, while the **item** column represents the movie's ID number corresponding to that user's rating. The **rating** column shows how much the user liked the movie, usually on a scale from 1 to 5. The **timestamp** column records when the rating was made, stored as a UNIX timestamp (seconds since January 1, 1970).

1. Load Dataset 'MovieLens.csv' and keep ratings of only those users who have at least 5 ratings.
2. Group each rating by userid and then arrange ratings from a specific user in increasing order of timestamp. The most recent rating should be the last entry. Form a test set by combining all most recent ratings from a user. All the other ratings are stored in train set. Recommendation models in this experiment will be built on train set and they will be tested on test set to validate their future prediction power.
3. Three Exploratory Data Analysis (EDA) steps need to be implemented. These will be used to analyze the results obtained. EDA 1: Print the range of movie ratings in order to ensure that dataset is not biased. EDA 2: Identify the most active users. EDA 3: Identify the most popular movies by count.
4. Model A (Popularity Model based on Bayesian mean score): Use the training set to identify Bayesian mean score for each movie. During test time the model just recommends the top-10 (based on this score) movies that user has not rated. $(\text{BayesMean}[i] = (C.m + n[i]m[i]) / (C + n[i])$ where $C=20$ (smoothing constant), m is the global average ratings across all movies, $n[i]$ is number of ratings for movie i and $m[i]$ is mean rating of movie i .
5. Make recommendations for the test set using Model A. Evaluate HitRate@10.
6. Model B (Item-Based Collaborative Filtering): Using the trainset create a user-item matrix(ui_mat) where rows are users and columns are movies. Each cell of this user-item matrix is the rating given by user j to movie k . To remove bias of the user, normalize each row by subtracting row mean. In the transformed matrix, replace all NA values with zero. Now calculate similarity index between columns by using correlation. The result is a square matrix whose (j,k) element indicate the similarity (in terms of rating) between movies. The values of this matrix lie in $[-1,1]$. The value +1 is highly correlated, 0 is no correlation and -1 is negatively correlated.
7. Exploratory: Make recommendations for the test set using Model B. Evaluate HitRate@10.

8. Exploratory: Implement a 2 sample Difference of proportion test for hitRate@10.