

Case Study

Netflix Recommendation System

Data Science Process

Step 1: **Business Understanding** - Netflix has thousands of movies and TV shows, so users often face difficulty deciding what to watch. The main business goal is to recommend and personalize movies/shows for each user based on their viewing history and ratings

. • Example: Suppose a user logs in but doesn't know what to watch. If Netflix suggests a trending action movie that matches their taste, they stay longer on the platform. But if they don't get good suggestions, they may stop using Netflix or switch to another service.

Step 2: **Data Understanding** - Netflix collects data from various sources. For this case study, we use the MovieLens dataset (public dataset) which has the following columns:

1. User ID → Example: 123 (identifies each user)
2. Movie ID → Example: M456 (identifies each movie)
3. Rating (1–5) → Example: 4 (shows how much user likes the movie)
4. Timestamp → Example: 1094789123 (when rating was given)
5. Movie Metadata → Example: Krish, 2004, Drama (title, release year, genres)

• Example: User 123 rated “Krish (2004, Drama)” as 4 stars on 10th Sept 2004. This tells us the user enjoys Drama movies.

Step 3: **Data Preparation** - Before building the recommendation models, the raw data needs to be cleaned and transformed to make it suitable for analysis:

1. Fix Missing Values - Some movies may have missing details like release year or genre. - Example: If the release year of a movie is missing, we can either fill it using external sources (like IMDb) or remove that movie from the dataset to avoid errors in the model.

2. Convert Timestamp

- Raw timestamps are in numerical format and not human-readable.
- Example: 1094789123 → converted to 2004-09-10 so that it's easier to analyze user behavior over time.

3. Encode Genres

- Textual genre data (like Drama, Action, Romance) is converted into numerical format using one-hot encoding.

- Example: • Drama = 1, Action = 0, Romance = 0 for a Drama movie. • “Krish (2004, Drama)” becomes [Drama=1, Action=0, Romance=0, ...].

Now data is ready for analysis and modeling.

Step 4: **Modeling** - Netflix uses two main approaches: 1. Collaborative Filtering - Finds patterns between users and movies using SVD (Singular Value Decomposition). –

Example: If users A and B have similar movie tastes, and user A watched a movie that B hasn't, that movie is recommended to B.

2. Content-Based Filtering

- Uses movie metadata like genres and release date.
- Example: If a user likes action movies, the system recommends another action movie to the user.

Step 5: **Evaluation** - After modeling, Netflix measures performance using RMSE, Precision, and Recall.

- Uses a hybrid approach: 80% Collaborative Filtering + 20% Content-Based Filtering.
- Example: If Netflix recommends 10 movies and the user watches 7 fully, the system has high precision, showing recommendations are relevant.

Step 6: **Deployment** - The final step is deployment. Users log in, and the system checks their personalized taste, ratings, and viewing history. The system is regularly updated with new ratings and user activity.

- Example: If a user starts watching K-Dramas, Netflix updates recommendations to show more Korean dramas.

This way, Netflix ensures every user gets a unique and engaging experience.