

Netflix Watching Time

Problem Statement:

Design model that can predict the Netflix watching time for a specific TV or movie for a particular user.

1. Problem overview

Challenges might be faced:

1. New users vs old users
2. Data sparse meaning some users might watch only a few titles of movie or tv
3. Temporary habit changing like switching from one genre or show or anything over time
4. External factors like some users might prefer watching on weekday or weekend and some time interval of the day like after work between 8 and 12 or other times.

2. Data Collection

Users data or Demographics:

Features: user_id, date_birth, gender, account_type, location, device_type, historical_watch_time per day, week, and month

Content (Movie, TV show and so on) Data:

Features: content_id, release_year, genre, rating, popularity_score (total watch count), duration

User-Content Interactions Data:

Features: start_time, end_time, stop_time, completion_rate, watch_frequency

Temporal Data:

Features: day_of_week, time_of_day (Morning, Afternoon, Evening), holiday

3. Data Preprocessing

1. Handling Missing data
2. Feature Engineering (creating new features based on user behavior)
3. Encoding categorical data
4. Scaling if needed
5. Cross Validation

4. Modeling

1. Linear Regression (baseline model)
2. Random Forest (if non-linearity)
3. Boosting Algorithms (for large and tabular data)

2. Model Evaluation

1. R2 score → getting in percentages how far predicting values from actual ones
2. MSE or MAE
3. RMSE → to get overall performance of a model
4. Residuals Testing