Guide to Building Recommendation Systems

Content-Based Filtering

CB systems create recommendations by capturing the similarity relationships between items based on their comments, descriptions, or attributes that users interact with.

1. Data Preparation

a. Data Cleaning:

Handle missing values (e.g., filling or removing them).

Analyze and correct anomalous data points (e.g., incorrect ratings).

b. Data Transformation:

Extract meaningful features from item descriptions or metadata.

Create a similarity matrix by comparing item attributes (e.g., using TF-IDF for text-based features).

2. Similarity Measurement

Calculate similarities between items using techniques such as:

Cosine Similarity: Measures the cosine of the angle between two vectors.

Jaccard Similarity: Compares the overlap between sets of attributes.

3. Recommendation Generation

a. Item Scoring:

Score items based on their similarity to items the user has interacted with.

b. Generate Recommendation List:

Rank items by their similarity scores and present the top-ranked items to the user.

4. Model Evaluation

Evaluation Metrics:

- **Accuracy**: MAE, RMSE.
- **Ranking Quality:** nDCG, ARHR, RHR.
- **User Experience:** Precision, Recall, F1-Score.
- **General Accuracy:** CHR (Catalog Hit Rate) evaluates the proportion of user-preferred items among all recommendations.
- **Exploration:** Serendipity, Novelty, Personalized Novelty.
- **Diversity:** Measure variability of recommendations.
- **Coverage:** Percentage of items that can be recommended.
- **Freshness:** Use **churn** to ensure dynamic and engaging recommendations.