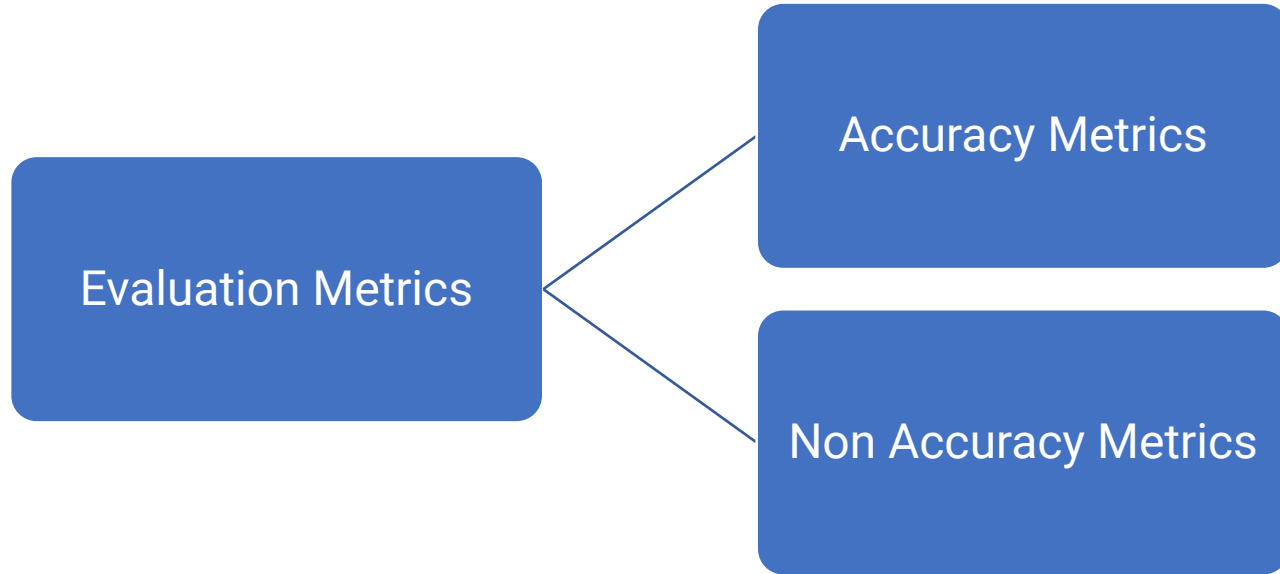


Evaluation of Recommender Systems

Why is evaluation difficult?

- Variability in datasets for different domains
 - Movie Recommender ($n_{\text{users}} \gg n_{\text{items}}$) vs Research Paper Recommender ($n_{\text{items}} \gg n_{\text{users}}$)
- Goals for evaluation may differ
 - Traditionally accuracy has been considered important
 - User Satisfaction is important – Difficult to measure
 - Coverage
 - Novelty
 - No of Purchases

Types of Evaluation Metrics



Accuracy Metrics: Types of Output

- Output types
 - **Prediction:** A (numerical) prediction indicating to what degree the current user will like or dislike a certain item
 - **Recommendation:** A top-N list of recommended items

Prediction

User	Movie	Predicted Rating
Tom	Argo	5
Tom	Seven	4
Tom	Righteous Kill	3

Top n Recommendations

Tom: {Argo, Seven, Righteous Kill}

Accuracy Metrics Taxonomy

