# **Pratilipi Recommendation System - Documentation**

## **Data Analysis**

### **User Interaction Dataset**

* **Columns**: user\_id, pratilipi\_id, read\_percentage, updated\_at
* **Purpose**: Tracks how much of each pratilipi a user has read, allowing us to infer preferences.

### **Metadata Dataset**

* **Columns**: author\_id, pratilipi\_id, category\_name, reading\_time, updated\_at, published\_at
* **Purpose**: Provides additional information about pratilipis for content-based recommendations.

### **Data Preprocessing**

* Merged both datasets using pratilipi\_id.
* Converted user\_id and pratilipi\_id into sequential integer mappings to optimize indexing.

## **Training Process**

1. **Collaborative Filtering (SVD Model)**:  
   * Used Surprise library’s SVD algorithm to model user reading behavior.
   * Trained on 75% of the dataset using train\_test\_split().
   * Predicted ratings for unseen pratilipis.
2. **Content-Based Filtering**:  
   * Converted category\_name into one-hot encoding.
   * Used cosine similarity to find similar pratilipis.
3. **Hybrid Model**:  
   * Combined SVD-based recommendations with content-based similar pratilipis.
   * Recommended a mix of user-based and content-based suggestions.

## **Model Choice**

### **Why SVD?**

* Handles sparse data efficiently.
* Captures latent factors in user preferences.

### **Why Cosine Similarity?**

* Provides intuitive content-based recommendations.
* Easily adaptable to metadata changes.

## **Summary**

The system predicts and recommends pratilipis based on user history and content similarities, leveraging a hybrid approach for better personalization.