Hard Path Pipeline

Hard Path: Hybrid Deep Learning with Sequential Patterns and Metadata Fusion

1. Data Input & Preprocessing:
- Load raw audio data with separated vocal and non-vocal tracks.
- Extract Mel-Spectrograms, STFT, and sequence mining features (e.g., Generalized Sequential Pattern (GSP)) from user-item interaction logs.
- Normalize features separately for vocal and non-vocal tracks.
- Encode categorical metadata (Artist, Genre, Key) as learned embeddings.

2. Feature Engineering:
- Develop multi-input feature sets combining spectrogram embeddings, sequential pattern features, and metadata embeddings.
- Use autoencoders or deep feature extractors for dimensionality reduction.
- Fuse vocal and non-vocal track features via learned representation layers.

3. Model Selection & Training:
- Use Nested Cross-Validation or Bayesian Optimization for hyperparameter tuning.
- Build hybrid deep learning architectures with CNNs or Transformers on spectrogram inputs and RNNs on sequential data.
- Combine with ensemble methods (XGBoost, stacking) on metadata and sequence features.

4. Model Evaluation:
- Evaluate with advanced metrics: Mean Average Precision (MAP), personalized recommendation accuracy, and user satisfaction metrics.
- Perform ablation studies on vocal vs non-vocal input contributions.
- Analyze error patterns per language, genre, and user segment.

Handling Vocal/Audio Tracks:
- Design separate model branches for vocal and non-vocal audio inputs.
- Augment vocal and audio tracks independently during training.
- Integrate multi-modal inputs with metadata embeddings for end-to-end training.

ML Techniques Used:
- Deep Learning (CNN, RNN, Transformers)
- Autoencoders
- Sequence Mining (GSP)
- XGBoost, Stacking
- Nested Cross-Validation
- Bayesian Optimization
- Mean Average Precision (MAP)
- Embeddings for categorical data
- Feature Normalization