

Easy Path Pipeline

Easy Path: Statistical Features from Vocal & Audio Tracks + Traditional ML

Pipeline Outline

1. Data Input & Preprocessing:

- Load audio files with separated vocal and non-vocal (instrumental/audio) tracks.
- Extract or load precomputed statistical features (Mean, Variance, Skewness, Kurtosis, RMS Energy, Zero Crossing Rate, Tempo, Loudness) separately for vocal and audio tracks.
- Normalize numerical features (e.g., Min-Max or Standard Scaling).
- Encode categorical metadata (if using, e.g., Genre, Artist).

2. Feature Engineering:

- Keep vocal track statistics as primary features.
- Optionally append audio track statistics as additional features.
- Combine all features into one feature vector per sample.

3. Model Selection & Training:

- Split dataset into train/test sets (Stratified K-Fold recommended).
- Train classical classifiers (Logistic Regression, Random Forest, or SVM) on combined feature vectors.

4. Model Evaluation:

- Evaluate with Accuracy, Precision, Recall, F1-score.
- Compare performance with vocal-only features vs vocal+audio features.
- Use confusion matrix to inspect language-specific errors.

Handling Vocal/Audio Tracks:

- Load and process vocal and audio tracks separately at every stage.
- Use vocal track features as main input; add audio features optionally.

Machine Learning Techniques Summary

ML Techniques Used:

- Logistic Regression
- Random Forest
- Support Vector Machines (SVM)
- Stratified K-Fold Cross-Validation
- Accuracy, Precision, Recall, F1-score metrics
- Confusion Matrix analysis
- Feature Normalization (Min-Max, Standard Scaling)
- Encoding categorical metadata (One-hot, Label Encoding)