AOTY Prediction Pipeline (Detailed)

Raw CSV

(all_albums_full.csv)

Schema Validation — *(Pandera)*

Cleaning Rules — *(nulls, dates)*

Min Ratings Filter — *(threshold=10)*

Audit Trail — *(data/audit/*.jsonl)*

Within-Artist Split — *(temporal ordering)*

Artist-Disjoint Split — *(new artist holdout)*

Split Manifests — *(hash verification)*

Train — *(N_train)*

Validation — *(N_val)*

Test — *(N_test)*

TemporalBlock — *(5 temporal)*

AlbumTypeBlock — *(one-hot)*

ArtistHistoryBlock — *(LOO stats)*

ArtistReputationBlock — *(prior mean)*

GenreBlock — *(PCA)*

CollaborationBlock — *(ordinal)*

FeaturePipeline — *(fit on train, transform all)*

PriorConfig — *(9 hyperparams)*

Hierarchical Model — *(artist effects)*

Non-centered Param — *(LocScaleReparam)*

Time-varying Effects — *(random walk)*

AR(1) Structure — *(stationary)*

MCMC Sampling — *(4 chains, GPU)* — *tune against*

R-hat — *(< 1.01)*

ESS — *(> 400)*

Divergences — *(= 0)*

LOO-CV — *(PSIS)*

Pareto-k — *(< 0.7)*

Calibration — *(coverage)*

Sensitivity — *(priors)*

Predictions — *(posterior mean)*

Uncertainty Bands — *(95% CI)*

Publication Artifacts — *(figures, tables)*

Legend:
- Data Input
- Sanitization
- Splitting
- Features
- Model
- Validation
- Output