Model Training Pipeline

LC_TRAIN Dataset

Model Definitions

1. Ridge (alpha)
2. Lasso (alpha)
3. PenalizedSplines (knots)
4. KNN (n_neighbors)
5. PenalizedLogNormal (alpha)

Pipeline Steps

1. Scaler (Standard/Robust/MinMax)
2. Feature Engineering:
   - vanilla (base)
   - interact_select (interaction terms + KBestSelect(f_regression)
   - pca_lda (new features: pca + lda + KBestSelect(f_regression)

Cross Validation

1. Split Strategy:
   - kfold (random)
   - rolling (time)
   - expanding (time)
2. Metric: RMSE/MAE/R^2

Model Selection

1. Average CV Performance
2. Standard Deviation
3. Best Parameters

Final Evaluation

1. Train Best Model
2. Evaluate on Holdout Set