

INPUT:
• 9 Datasets
• 159,811 Samples
• Multi-domain

Data Collection
(9 Datasets)

Data Preprocessing & Quality Control

Data Validation & Leakage Detection

Baseline Models
(LASSO, RIDGE, MEAN)

Traditional ML
(RF, XGB, SVR)

Deep Learning
(LSTM, Transformer)

MODELS:
• Baseline (3)
• Traditional ML (3)
• Deep Learning (2)

Hyperparameter Optimization & Cross-Validation

Performance Evaluation

Statistical Significance

Robustness Analysis

Results Integration & Analysis

METRICS:
• R² Score
• Statistical Test
• Robustness

Cross-Dataset Comparison

Model Selection Guidelines

FINDINGS:
• Data Quality > Quantity
• Model Choice Matters
• Validation Critical

Best Practices & Recommendations

KEY RESULTS:
• RF: R²=0.855 (Best)
• 7/9 Datasets Validated
• LSTM: 3/7 Best Performance