Model Comparison Summary - Novartis Datathon 2025

Generated: 2025-11-29

Summary Table

Model	Scenario	Official Metric	RMSE	MAE	Training Time (s)	Notes
CatBoost	1	0.6853	0.2472	0.1794	14.06	Best for S1
CatBoost	2	0.2413	0.1982	0.1229	14.37	Best for S2
Linear	1	0.7619	0.2618	0.1969	0.02	Fast baseline
Linear	2	0.3104	0.2096	0.1343	0.02	Fast baseline
Historical Curve	1	0.8195	0.2658	0.2069	0.20	Simple baseline
Historical Curve	2	0.4037	0.2266	0.1526	0.21	Simple baseline
Neural Network	1	0.9648	0.2999	0.2329	119.59	High variance
Neural Network	2	0.5815	0.2884	0.2015	238.34	High variance
Hybrid (CatBoost)	1	0.7885	0.2509	0.1815	6.06	Physics+ML
Hybrid (CatBoost)	2	0.2773	0.2049	0.1265	6.85	Physics+ML
ARIHOW	1	1.2069	0.6722	0.6026	68.81	Time series (poor fit)
ARIHOW	2	0.8350	0.6306	0.5549	66.41	Time series (poor fit)
LightGBM	1	SEGFAULT	-	-	-	Apple Silicon issue
LightGBM	2	SEGFAULT	-	-	-	Apple Silicon issue
XGBoost	1	SEGFAULT	-	-	-	Apple Silicon issue
XGBoost	2	SEGFAULT	-	-	-	Apple Silicon issue

Rankings by Official Metric

Scenario 1 (Lower is better)

1. **CatBoost**: 0.6853 ⭐ BEST
2. Linear: 0.7619
3. Hybrid: 0.7885
4. Historical Curve: 0.8195
5. Neural Network: 0.9648
6. ARIHOW: 1.2069

Scenario 2 (Lower is better)

1. **CatBoost**: 0.2413 ⭐ BEST
2. Hybrid: 0.2773
3. Linear: 0.3104
4. Historical Curve: 0.4037
5. Neural Network: 0.5815
6. ARIHOW: 0.8350

Key Insights

1. **CatBoost dominates** both scenarios with the lowest official metric scores.
2. **Hybrid (Physics+ML)** performs well, especially in Scenario 2, combining interpretability with good performance.
3. **Linear model** provides a strong baseline with minimal training time.
4. **Historical Curve** is a simple yet reasonable baseline.
5. **Neural Network** shows high variance and overfitting tendencies.
6. **ARIHOW** (ARIMA + Holt-Winters) performs poorly due to limited historical data and poor generalization.
7. **LightGBM/XGBoost** crash on Apple Silicon (M-series) - need to run on Linux/Intel for comparison.

Recommendations

1. **For production**: Use CatBoost as primary model
2. **For interpretability**: Consider Hybrid model (physics baseline + ML residuals)
3. **For ensemble**: Blend CatBoost + Hybrid + Linear
4. **For fast inference**: Linear model provides reasonable accuracy

Technical Notes

- All models trained with 80/20 stratified split by brand/bucket
- Sample weights applied based on bucket distribution
- Features cached for consistent comparison
- LightGBM/XGBoost require different environment (segfault on Apple Silicon M-series)