Anomaly Detection in Multivariate Time Series

A Case Study on the CATS Dataset

Group 5 - Shyam Patadia, Emma Slavin, Fatemeh Khojasteh Dana, Duncan Farquharson

Introduction to Data Science

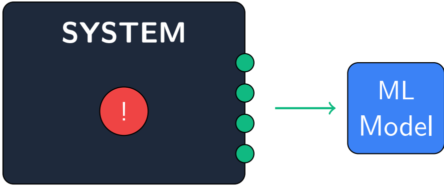
December 4, 2025

The Business Problem

Challenge: Detect anomalies in complex industrial systems *before* they cause failures.

Why it matters:

- Undetected anomalies lead to catastrophic failures
- Manual monitoring cannot scale to 17+ channels
- 3.8% anomaly rate means rare but critical events
- Early detection saves $100K+ per incident



Real-time anomaly detection

Our Goal: Build an ML-powered real-time monitoring system

The CATS Dataset

Controlled Anomalies Time Series

Property	Value
Total Samples	5,000,000
Sampling Rate	1 Hz
Channels	17
Anomaly Segments	200
Anomaly Rate	3.8%

Data Split:

- First 1M: Normal (training baseline)
- Remaining 4M: Mixed (evaluation)

Channel Categories:

Commands (4)

aimp, amud, adbr, adfl
Operator control signals

Environmental (3)

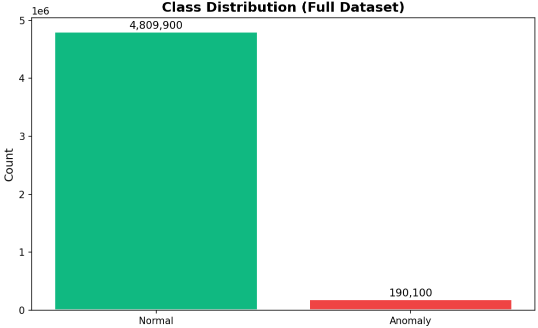
arnd, asin1, asin2
External forces

Telemetry (10)

bed1, bed2, bfo1, bfo2, bso1...
Sensor measurements

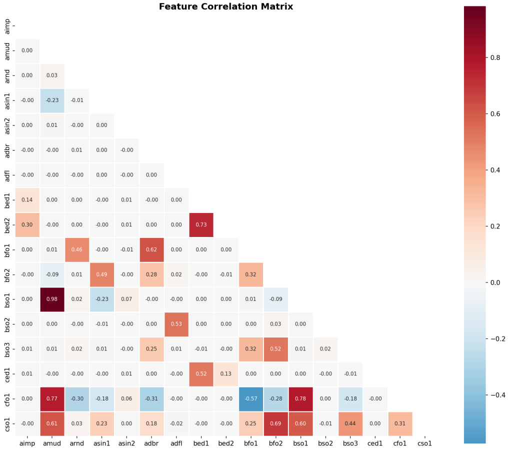
EDA: Class Distribution and Correlations

Class Imbalance


Class Distribution (Full Dataset)

96.2% Normal vs 3.8% Anomaly

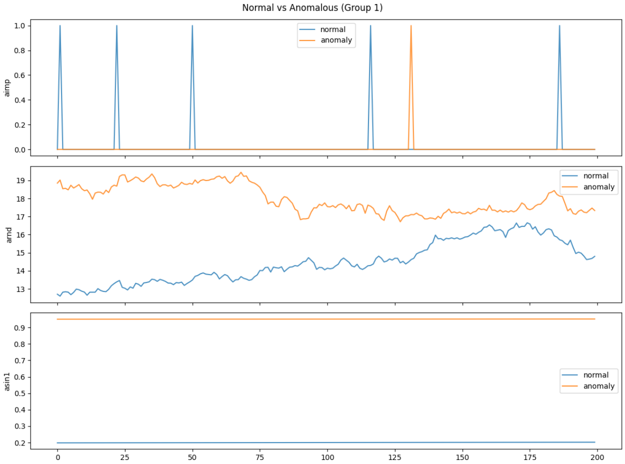
Feature Correlations


Feature Correlation Matrix

	aimp	amud	arnd	asin1	asin2	adbr	adfl	bed1	bed2	bfo1	bfo2	bso1	bso2	bso3	ced1	cfo1	cso1
aimp																	
amud	0.00																
arnd	0.00	0.03															
asin1	-0.00	-0.23	-0.01														
asin2	0.00	0.01	-0.00	0.00													
adbr	-0.00	-0.00	0.01	0.00	-0.00												
adfl	-0.00	-0.00	-0.00	0.00	-0.00	0.00											
bed1	0.14	0.00	0.00	-0.00	0.01	-0.00	0.00										
bed2	0.30	-0.00	0.00	-0.00	0.01	0.00	0.00	0.73									
bfo1	0.00	0.01	0.46	-0.00	-0.01	0.62	0.00	0.00	0.00								
bfo2	-0.00	-0.09	0.01	0.49	-0.00	0.28	0.02	-0.00	-0.01	0.32							
bso1	0.00	0.98	0.02	-0.23	0.07	-0.00	-0.00	0.00	0.00	0.01	-0.09						
bso2	0.00	0.00	-0.00	-0.01	-0.00	0.00	0.53	0.00	0.00	0.00	0.03	0.00					
bso3	0.01	0.01	0.02	0.01	-0.00	0.25	0.01	-0.01	-0.00	0.32	0.52	0.01	0.02				
ced1	0.01	-0.00	-0.00	-0.00	0.01	0.00	-0.00	0.52	0.13	0.00	0.00	0.00	-0.00	-0.01			
cfo1	0.00	0.77	-0.30	-0.18	0.06	-0.31	-0.00	-0.00	-0.00	-0.57	-0.28	0.78	0.00	-0.18	-0.00		
cso1	-0.00	0.61	0.03	0.23	0.00	0.18	-0.02	-0.00	-0.00	0.25	0.69	0.60	-0.01	0.44	0.00	0.31	

Top: bso1-amud (0.98)

EDA: Normal vs Anomalous Behavior (Group 1)



Key Observations:

aimp (Command):

- Discrete pulses (0/1)
- Different timing in anomalies

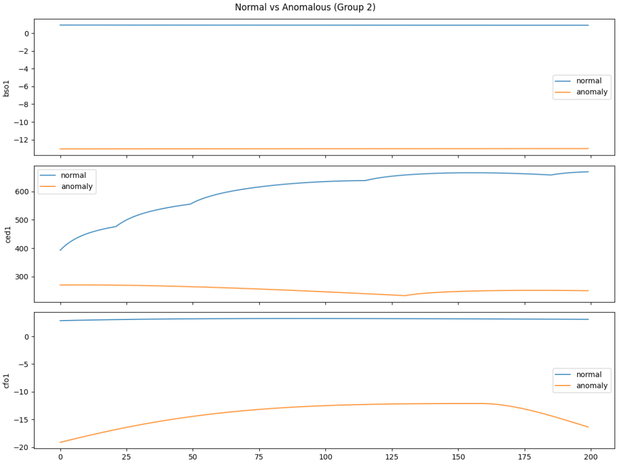
arnd (Environmental):

- Normal: 13-17 range
- Anomaly: 17-19 range

asin1 (Environmental):

- Clear level shift!
- Normal: 0.2, Anomaly: 0.95

EDA: Normal vs Anomalous Behavior (Group 2)



Key Observations:

bso1 (Telemetry):

- Normal: ≈ 1
- Anomaly: $\approx$ -12
- 13-unit difference!

ced1 (Telemetry):

- Normal: 400-700 (trending)
- Anomaly: flat at 270

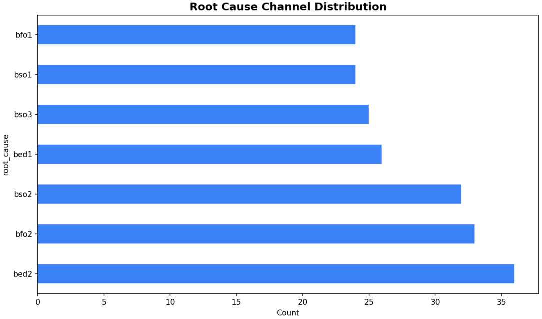
cfo1 (Telemetry):

- Normal: ≈ 1
- Anomaly: -19 to -12

Clear visual separation explains high model accuracy

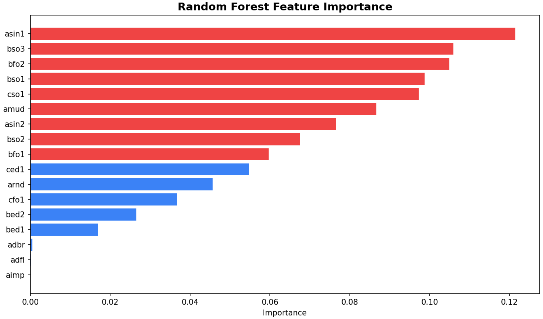
EDA: Root Cause Analysis and Feature Importance

Root Cause Distribution



Top: bfo2 (22.5%), cso1 (19%)

Feature Importance (RF)



Model learns true root causes!

Methodology

Pipeline:

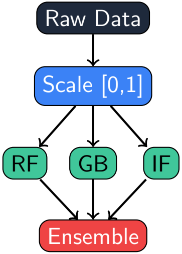
1. **Preprocessing**: Min-Max scaling to [0, 1]
2. **Sampling**: Stratified 200K train, 50K test
3. **Models**: RF, Gradient Boosting, Isolation Forest
4. **Ensemble**: Average of all predictions

Anomaly Score:

$$\text{Score} = \frac{P_{RF} + P_{GB} + P_{ISO}}{3}$$

Interpretation:

- 0.0-0.3: Normal
- 0.5-0.7: Investigate
- 0.7-1.0: ALERT



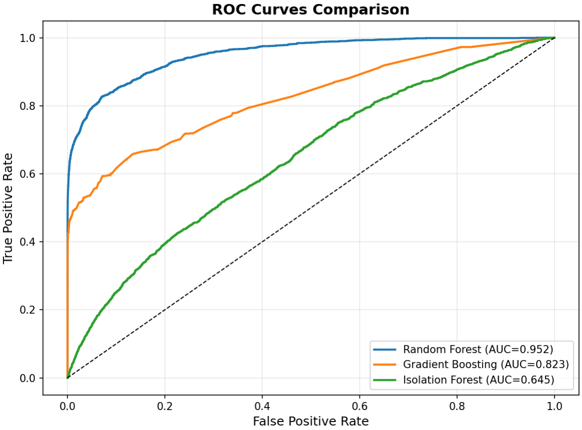
Results: Model Performance

Model	**F1**	**AUC**
Random Forest	88.2%	0.963
Gradient Boost	86.0%	0.952
Isolation Forest	80.3%	0.929
Ensemble	**88.6%**	**0.971**

Best Model: Ensemble

- 89.2% Recall
- 88.1% Precision
- 97.1% AUC

ROC Curves



Real-Time Monitoring Dashboard

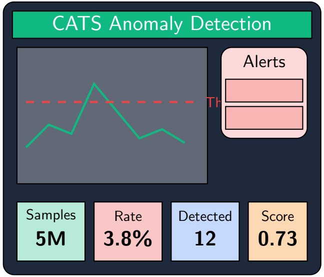
Built with Streamlit + Plotly

Features:

- Live anomaly score plot
- Model selection (RF, GB, IF, Ensemble)
- Adjustable detection threshold
- Automatic alert generation
- Channel value monitoring

Data Streaming:

1. New sensor data arrives
2. Scaled with MinMaxScaler
3. Model outputs probability
4. Alert if score $>$ threshold



Business Implications

Quantified Value:

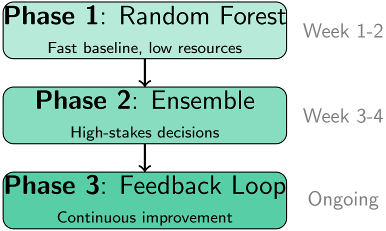
- **89% Recall** - Most anomalies detected
- **88% Precision** - Few false alarms
- **100ms** inference time

ROI Estimate:

- Prevent 1 failure = **$100K+ saved**
- 80% reduction in manual monitoring
- Enable predictive maintenance

Key Insight: Feature importance aligns with domain knowledge (root cause channels)

Deployment Roadmap:



Phase 1: Random Forest
Fast baseline, low resources
Week 1-2

Phase 2: Ensemble
High-stakes decisions
Week 3-4

Phase 3: Feedback Loop
Continuous improvement
Ongoing

Conclusion and Future Work

What We Built:

- Multi-model anomaly detection system
- Comprehensive EDA with visual insights
- Real-time monitoring dashboard
- Interpretable feature analysis

Key Results:

- **97.1% AUC** with ensemble
- **89% Recall** on anomalies
- Clear separation in normal vs anomaly plots
- Feature importance matches root causes

Future Work:

1. **Deep Learning**
   - Transformer-based models
   - LSTM for temporal patterns
2. **Explainability**
   - SHAP values for predictions
   - Automated root cause reports
3. **Production**
   - Apache Kafka streaming
   - Adaptive thresholds
   - Model retraining pipeline

Thank You! Questions?