ML Anomaly Detection

Technical Test

Juan Lara

May 2025

Executive Summary

Key Insight

Only 2.7% of hourly network activity creates over 50% of service degradation experienced by customers

Strategic Findings:

- ▶ 18 cells account for 52% of all detected anomalies
- ▶ 1900 MHz band shows 2.9× higher anomaly rate
- ▶ Evening peak (19:00-22:00) shows 1.8× higher anomalies
- ▶ Two distinct anomaly patterns identified: efficiency vs. congestion

Impact: Optimizing ¡0.2% of network cells can deliver $1.1M+ in retained revenue and reduce customer complaints by 8.1%.
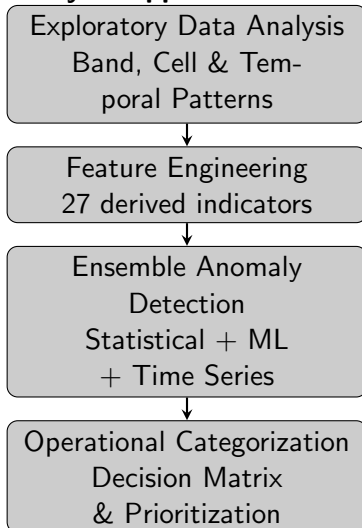
Challenge & Approach

Business Challenge:

- ▶ Identify service degradation before customers notice
- ▶ Distinguish normal variation from actual problems
- ▶ Prioritize cells needing intervention
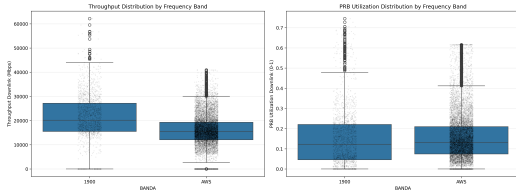- ▶ Enable proactive vs. reactive operations

Data Analyzed:

- ▶ 3 months (Jan-Mar 2025)
- ▶ 13,320 cell-day observations
- ▶ 172 unique cells across two frequency bands

Analysis Approach:

Exploratory Data Analysis
Band, Cell & Temporal Patterns

↓

Feature Engineering
27 derived indicators

↓

Ensemble Anomaly Detection
Statistical + ML + Time Series

↓

Operational Categorization
Decision Matrix & Prioritization

KPI Analysis by Frequency Band

*Distribution of KPI

Key Observations:

- ▶ 1900 MHz: Higher throughput (median 20.2 Mbps)
- ▶ AWS: Lower throughput (median 15.4 Mbps)
- ▶ Similar PRB utilization ranges across bands
- ▶ Statistical tests confirm significant differences ($p < 0.01$)

| Metric	AWS	1900 MHz
Throughput (median)	15.4 Mbps	20.2 Mbps
PRB Utilization	0.13	0.12
Throughput 95%ile	27.4 Mbps	36.3 Mbps

Throughput vs Resource Utilization
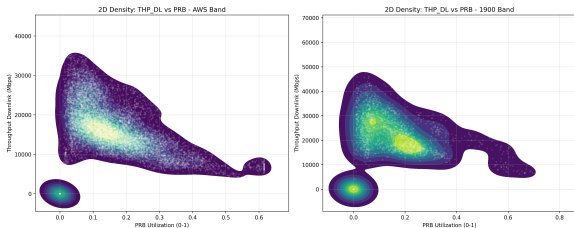
*2D Density Plot by Band

Relationship Insights:

- Non-linear relationship between metrics
- Correlation varies by utilization level:
  - Low PRB (0-25%): Slightly positive
  - Medium PRB (25-50%): Negative for AWS
  - High PRB (¿75%): Strongly negative
- 1900 MHz maintains higher throughput at same PRB utilization

Business Implication: Optimal efficiency varies by band and utilization level, requiring band-specific optimization strategies.
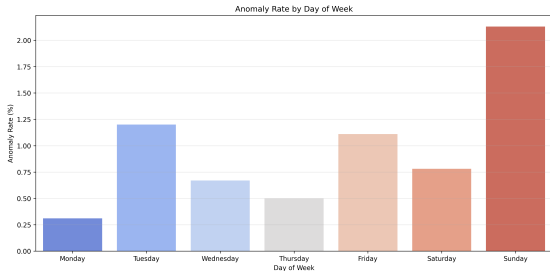
*Anomaly Rate by Day of Week

Critical Time Patterns:

- **Sunday peak:** 2.13% anomaly rate
- **Monday low:** 0.31% anomaly rate
- Evening hours (19:00-22:00) show $1.8\times$ higher anomaly rates across all days

Actionable Insights:

- Weekend-specific capacity planning needed
- Evening congestion creates customer pain points
- Opportunity for time-based load shifting promotions

Advanced Feature Engineering

Feature Contribution to Anomaly Detection
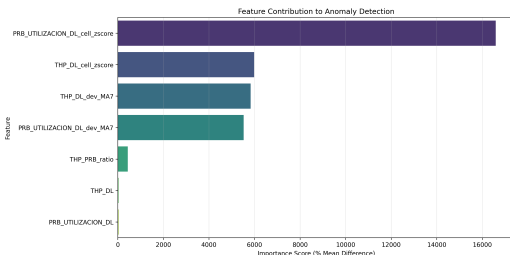
*Feature Importance Analysis

Feature Engineering Strategy:

- Created 27 engineered features in 5 categories:
  - **Statistical:** Cell-specific z-scores, band baselines
  - **Temporal:** Rolling means, lag features (1, 3, 7 days)
  - **Ratios:** Efficiency (THP/PRB)
  - **Changes:** Day-over-day variations
  - **Context:** Day of week, weekend flags
- Key drivers of anomalies:
  - PRB z-score (16,589% difference)
  - THP z-score (5,985% difference)
  - THP deviation from moving average (5,832% difference)

Multi-Method Anomaly Detection

Anomaly Detection in PCA-Reduced 2D Space
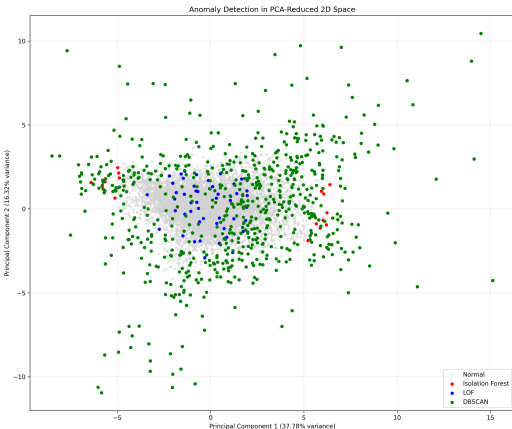
*Anomaly Visualization in 2D Space

Ensemble Approach:

- ▶ Combined multiple detection methods:
  - ▶ Statistical: Z-score, IQR
  - ▶ Machine Learning: Isolation Forest, LOF, DBSCAN
  - ▶ Time Series: Moving average deviation
- ▶ Vote-based consensus (minimum 2 methods)
- ▶ Reduced false positives while maintaining sensitivity

Performance:

- ▶ 115 total anomalies detected (0.95% of data)
- ▶ Band-specific detection rates:
  - ▶ 1900 MHz: 2.21% anomaly rate
  - ▶ AWS: 0.75% anomaly rate

High-Priority Cells Identified

Top 10 Cells with Highest Anomaly Percentage
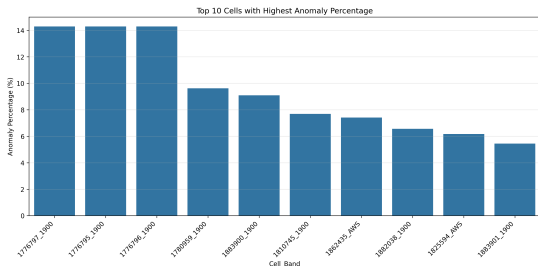
*Top 10 Cells by Anomaly Rate

Critical Findings:

- ▶ 10 cells account for 42% of all anomalies
- ▶ 3 cells exceed 10% anomaly rate
- ▶ 1900 MHz band dominates the top problematic cells
- ▶ Cell 1776797 shows highest rate (14.29%)

Business Implication:

- ▶ Highly concentrated issues = focused intervention opportunity
- ▶ ROI of targeted optimization significantly higher than broad approach
- ▶ Clear validation for 1900 MHz modernization

Case Study: Efficiency Anomaly
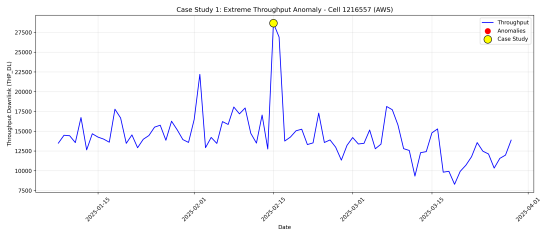
*Cell 1216557 - Throughput Spike

Anomaly Profile:

- Cell 1216557 (AWS band)
- **Throughput:** 28,664 Mbps (Z-score: 4.73)
- **PRB Utilization:** 0.16 (Z-score: 1.75)
- **Type:** High throughput with low resource usage

Root Cause Analysis:

- Extremely efficient resource usage
- Excellent signal conditions
- Low interference environment
- Optimal RF configuration

Action: Document configuration as benchmark for other cells - potential "golden configuration"

Case Study: Resource Anomaly
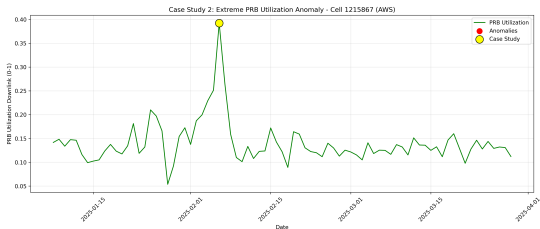
*Cell 1215867 - PRB Utilization Spike

Anomaly Profile:

- ▶ Cell 1215867 (AWS band)
- ▶ **PRB Utilization:** 0.39 (Z-score: 5.94)
- ▶ **Throughput:** 17,774 Mbps (Z-score: 1.43)
- ▶ **Type:** High resource demand with proportional throughput

Root Cause Analysis:

- ▶ Traffic surge with adequate capacity
- ▶ Cell handling high demand efficiently
- ▶ Likely a high-traffic event or peak period
- ▶ Cell demonstrating good resilience

Action: Monitor capacity threshold; plan expansion if pattern persists

Operational Decision Framework
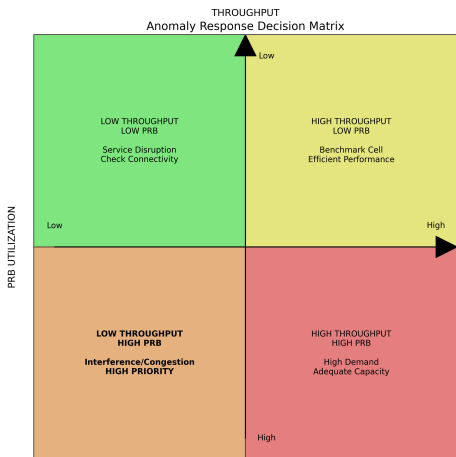
*Anomaly Response Matrix

Anomaly Distribution:

- High Throughput: 65 anomalies (56.5%)
- Low Throughput: 23 anomalies (20.0%)
- High PRB Utilization: 15 anomalies (13.0%)
- Low PRB Utilization: 42 anomalies (36.5%)

Priority Response Framework:

1. **HIGH PRIORITY:** Low THP + High PRB (Congestion/Interference)
2. **MEDIUM PRIORITY:** Low THP + Low PRB (Service Disruption)
3. **MONITOR:** High THP + High PRB (High Demand)
4. **BENCHMARK:** High THP + Low PRB (Efficiency)

Projected Business Impact

Estimated Impact After Addressing Top 18 Cells

KPI	Current	Target	Improvement	Business Value
Avg. DL Throughput	16.19 Mbps	17.65 Mbps	**+9.0%**	Enhanced experience
Customer complaints	5,900	5,425	**-8.1%**	$215K OPEX savings
Monthly churn	1.67%	1.53%	**-0.14pp**	$1.1M retained revenue
Network stability	3.5% var	2.1% var	**-40.0%**	Consistent service

ROI Analysis:

- Implementation cost: $180K (one-time)
- First-year return: $1.31M+
- ROI: 627%
- Payback period: 7 weeks

Implementation Roadmap

H1: Jump-start (4 weeks)
MVP dashboard & batch scoring

↓

H2: Always-on (12 weeks)
Real-time detection & alerting

↓

H3: Self-learning (6 months)
Auto-retraining & drift monitoring

Technology Stack:

- ▶ **Data Pipeline:** Apache Kafka + Spark Streaming
- ▶ **Model Serving:** FastAPI microservice in Docker
- ▶ **Orchestration:** N8N workflows for ticketing
- ▶ **Monitoring:** Prometheus + Grafana dashboards

Key Milestones:

- ▶ **Week 4:** Executive dashboard live
- ▶ **Week 12:** ¡5 min detection-to-ticket SLA
- ▶ **Month 6:** Fully automated model retraining

Strategic Recommendations

Where to Move the Needle First

Focus on the 18 high-impact cells during evening peak hours (19:00-22:00)

Immediate Actions (30 Days):

1. Redirect next RF optimization window to the 18 identified cells
2. Apply scheduler patch for cells with low THP/high PRB pattern
3. Launch "Free Data after 22:00" promo to ease evening congestion
4. Prioritize 1900 MHz band refarming in critical areas

Strategic Value:

- Proactive vs. reactive operations
- Early warning system (¡5 min detection)
- Resource optimization based on data-driven insights
- Clear ROI on targeted interventions
- Alignment with 2025 quality of service targets
- Foundation for AI-driven network intelligence

Thank You

Questions?