

# ML Insights Summary

## Device Failure Prediction (Classification)

Team 3: Afreen Alam, Sydney Dao, Manjari Kannan, Veena Komatineni, Shreya Tarat

### 1. Problem Definition

**Goal:** Predict if a smart home IoT device will fail (go offline)

**Target Variable:** device\_failed - Binary classification (0 = Active, 1 = Failed/Offline)

**Business Value:** Proactive device maintenance enables the smart home company to reduce customer complaints, improve satisfaction through proactive outreach, extend device lifespan, and optimize inventory management.

### 2. Model Performance Summary

Model	Accuracy	AUC-ROC	F1-Score
Logistic Regression (Best)	0.7946	<b>0.8136</b>	0.5476
Random Forest	0.7189	0.7669	0.3158
Gradient Boosting	0.6865	0.7237	0.3556

**Best Model:** Logistic Regression achieves the highest AUC-ROC score (0.8136), indicating strong discriminatory power between failed and active devices. The interpretable coefficients also provide clear insights into risk factors.

### 3. Key Predictive Features

Based on Logistic Regression coefficients from the best-performing model:

Rank	Feature	Coefficient	Interpretation
1	home_offline_rate	<b>+1.0446</b>	↑ Higher home offline rate strongly increases failure risk
2	recent_avg_value_7d	<b>-0.5291</b>	↓ Higher recent activity DECREASES failure risk (protective)
3	variance_numeric	<b>+0.5027</b>	↑ High reading variability increases failure risk
4	offline_devices_in_home	<b>+0.1619</b>	↑ More offline devices in home increases risk
5	device_age_days	<b>+0.0967</b>	↑ Older devices have slightly higher risk

**Note:** Positive coefficients increase failure probability; negative coefficients decrease failure probability (protective factors).

### 4. Key Insights

#### Insight 1: Home-Level Risk Clustering (Strongest Factor)

The home\_offline\_rate coefficient (+1.04) is by far the strongest predictor. Devices in homes with existing offline devices are significantly more likely to fail. This suggests environmental factors (Wi-Fi issues, power problems) or systematic issues affect multiple devices in the same home.

- Recommendation:** When one device fails, proactively check ALL devices in that home. Prioritize homes with >50% offline rate for immediate support outreach.

#### Insight 2: Recent Activity is Protective (Key Finding)

The recent\_avg\_value\_7d coefficient (-0.53) is negative, meaning higher recent activity DECREASES failure risk. Devices that are consistently active and generating readings are healthier. Conversely, devices with declining or low recent activity are at higher risk of failure.

- Recommendation:** Flag devices with declining activity or no readings in the past 24-48 hours. These are early warning signs of impending failure.

### **Insight 3: Reading Variability Signals Instability**

The variance\_numeric coefficient (+0.50) indicates that devices with highly variable sensor readings are more likely to fail. Unstable readings may indicate hardware degradation, connectivity issues, or environmental interference.

- **Recommendation:** Monitor reading variance and flag devices with abnormally high variability for diagnostic checks.

### **Insight 4: Device Age Has Moderate Impact**

While device\_age\_days has a positive coefficient (+0.10), it is not among the top predictors. Device age contributes to failure risk but is less important than home-level factors and activity patterns.

- **Recommendation:** Consider device age as a secondary factor. Focus primary attention on home offline rates and activity monitoring.

## **5. Business Recommendations**

### **Immediate Actions**

1. Contact customers with 2+ offline devices immediately (high home\_offline\_rate)
2. Create automated alerts for devices with no readings in past 24 hours
3. Flag devices showing abnormally high reading variance

### **Medium-Term Initiatives**

1. Deploy the ML model to score all devices daily based on failure probability
2. Build real-time activity monitoring dashboard tracking recent\_avg\_value trends
3. Implement "home health score" based on offline rate for customer success prioritization

### **Long-Term Strategy**

1. Investigate root causes of home-level failures (network issues, power quality)
2. Develop predictive maintenance workflows triggered by activity decline
3. Collect temporal failure data to build true failure\_within\_7\_days predictions