ML Insights Summary

Device Failure Prediction (Classification)

Team 3: Afreen Alam, Sydney Dao, Manjari Kannan, Harshaveena 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	**0.8136**	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**	**+1.0446**	↑ Higher home offline rate strongly increases failure risk
2	**recent_avg_value_7d**	**-0.5291**	↓ Higher recent activity DECREASES failure risk (protective)
3	**variance_numeric**	**+0.5027**	↑ High reading variability increases failure risk
4	**offline_devices_in_home**	**+0.1619**	↑ More offline devices in home increases risk
5	**device_age_days**	**+0.0967**	↑ 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 such as Wi-Fi issues or power problems, or systematic issues that affect multiple devices in the same home. When one device fails, it is recommended to proactively check all devices in that home. Homes with greater than 50% offline rate should be prioritized 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. To address this, devices with declining activity or no readings in the past 24-48 hours should be flagged, as 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. Reading variance should be monitored and devices with abnormally high variability should be flagged 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. Therefore, device age should be considered as a secondary factor, with primary attention focused on home offline rates and activity monitoring.

5. Business Recommendations

Immediate Actions

The most urgent priority is to contact customers with two or more offline devices immediately, as these homes show high offline rates indicating systemic issues. Automated alerts should be created for devices with no readings in the past 24 hours to catch early warning signs of failure. Additionally, devices showing abnormally high reading variance should be flagged for proactive investigation.

Medium-Term Initiatives

To build on immediate actions, the ML model should be deployed to score all devices daily based on failure probability, enabling systematic monitoring across the device fleet. A real-time activity monitoring dashboard tracking recent average value trends will help identify declining device health before complete failure occurs. Furthermore, implementing a "home health score" based on offline rate will allow customer success teams to prioritize their outreach efforts effectively.

Long-Term Strategy

For sustained improvement, the company should investigate root causes of home-level failures, such as network infrastructure issues or power quality problems, to address underlying environmental factors. Developing predictive maintenance workflows triggered by activity decline will enable truly proactive interventions. Finally, collecting temporal failure data will support building enhanced models that predict failure within specific time windows, such as failure_within_7_days, providing even more actionable insights for the business.