PumpSureAI: Predictive Maintenance with Sensor Data

Objective: To identify pump failures before they occur using advanced analytics.

Outcome: Minimized downtime and optimized maintenance scheduling.

Approach: Data-driven insights and machine learning models.

Data Overview and Preprocessing

Initial Inspection: Dataset structure, types, missing values.

Missing Values: Imputed based on analysis.

Outliers: Quantile-based thresholds applied.

Correlation Analysis: High-correlation features identified and visualized.

Advanced Feature Engineering

1. Critical Sensor Selection:
   - Correlation threshold (>0.3).

2. Derived Features:
   - Rolling windows: Mean, median, standard deviation.
   - Lagged data: 1, 5, 10-minute lags.
   - Differencing for rate-of-change.
   - Z-scores for anomaly detection.

3. Temporal Features:
   - Time-based (day, hour, weekday).
   - Event-based (time-to-failure).

Model Performance Summary

Models Explored:

1. Logistic Regression (LR): Test F1 = 0.9473.
2. AdaBoost: Test F1 = 0.9914.
3. GBM: Test F1 = 0.9698.
4. XGBoost: Test F1 = 0.9806.

- Deep Learning: Explored but not superior to ensemble models.
- Best Model: AdaBoost - robust and balanced.

Key Data Visualizations

1. Correlation Heatmap: Highlights redundancy and selection.
2. Sensor Trends: Rolling averages for pattern analysis.
3. Anomaly Detection: Z-score-based visualization.
4. Model Comparison: Training vs. Testing F1 scores (bar chart).

Detailed Model Results

	precision	recall	f1-score	support
0	1.00	0.99	0.99	32186
1	0.98	1.00	0.99	14484
accuracy			0.99	46670
macro avg	0.99	0.99	0.99	46670
weighted avg	0.99	0.99	0.99	46670

Final Insights and Recommendations

- Achievements:
- Comprehensive pipeline for predictive maintenance.
- Ensemble models outperform alternatives.

Next Steps:

1. Deploy AdaBoost model.
2. Real-time sensor data integration.
3. Explore advanced deep learning techniques for improvement.