

Forecasting & Anomaly Detection Pipeline

Candidate: Kowsalya P

Project Link: <https://github.com/kowsalya-42/Multi-Site-Operational-Forecasting-and-Anomaly-Detection-Pipeline.git>

Email id: kowsalyap.ai@gmail.com & kowsalyaparama@gmail.com

Date: September 10, 2025

Project Objective

Develop a scalable pipeline to forecast 14-day site-level production (units_produced) and power consumption (power_kwh), and detect operational downtime anomalies using interpretable models.

Data & Methodology

- Utilized daily operational data (production, power, downtime) merged with site metadata (region, commissioning year).
- Engineered temporal features (day of week, month, lags, rolling averages) and applied categorical encoding.
- Built baseline (linear regression) and improved models (XGBoost) for robust forecasting.
- Employed rolling-window z-score method for anomaly detection, balancing sensitivity and interpretability.

Results

- XGBoost models delivered superior accuracy over baselines, with reduced MAE and MAPE metrics.
- Anomalies flagged corresponded tightly with known downtime incidents, aiding early detection.
- Regional and operational factors demonstrated significant impact on performance metrics.

Implementation & Deliverables

- Modular codebase structured into data loading, feature engineering, modeling, and anomaly detection modules.
- CLI tool developed for flexible forecasting and alert generation by site and date range.
- Outputs saved as CSV files: forecast_units.csv, forecast_power.csv, and alerts.csv.

Business Impact & Future Work

- Enables proactive operational planning and maintenance scheduling to minimize downtime.
- Designed for automation and client integration with clear, explainable outcomes.
- Recommendations include incorporating external factors (weather, holidays) and building client-facing visual dashboards.