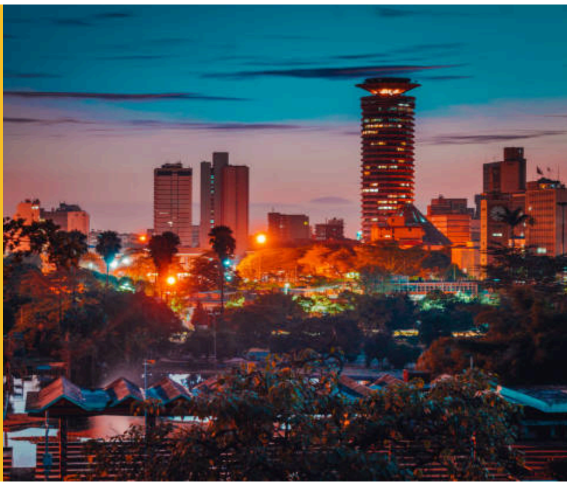
AI PREDICTIVE FRAMEWORK

HYBRID SOLAR DIESEL SYSTEM



Executive Summary & Vision

The predictive analytics framework transforms a commercial centre's hybrid solar system into a smart, cost-efficient energy solution, tailored to Kenya's unique challenges and opportunities. This in-progress implementation framework aims to deliver a fully operational system that forecasts solar production, predicts building demand, and optimizes battery and grid use — slashing energy costs by 10-20% and ensuring reliable power despite outages. For a shopping complex, office building, or small factory, this means turning abundant sunlight into predictable savings, reducing reliance on Kenya Power's pricey grid, and boosting sustainability credentials. Built on lightweight models and affordable hardware, it sidesteps Kenya's GPU scarcity, making it practical for local adoption. This isn't just a technical upgrade — it's a competitive edge, positioning the centre as a leader in Kenya's renewable energy push under Vision 2030, with a scalable blueprint for future growth.

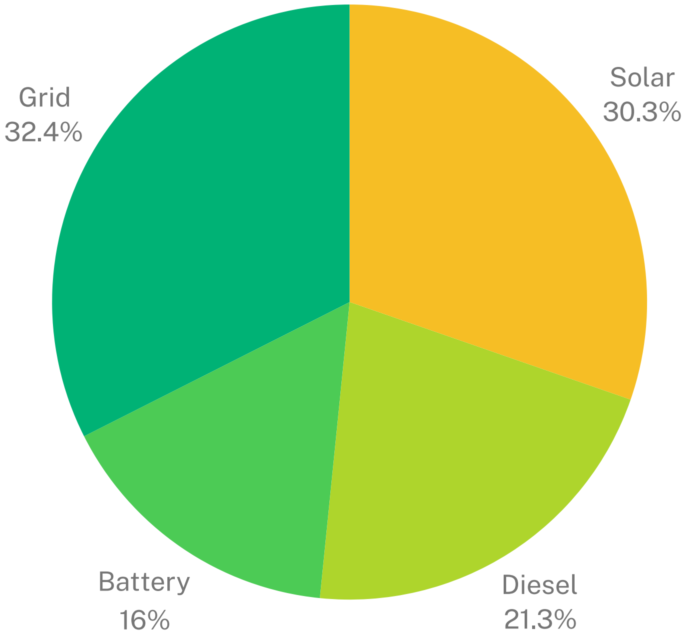


Figure One | Top Energy Sources for Commercial settings

Project Objectives

The predictive analytics framework is designed to unlock the full potential of a commercial centre's hybrid solar system, addressing its energy challenges with precision and efficiency. Tailored to a single site — be it a Nairobi office block or a Mombasa factory — the objectives deliver measurable value: reliable power, lower costs, and smarter resource use. By forecasting energy production and demand, optimizing battery and grid decisions, and ensuring scalability, the system turns raw data into actionable insights. It's built to thrive in Kenya's sunny yet unpredictable climate, using affordable tools to meet operational needs and stakeholder expectations, all while aligning with the centre's financial and sustainability goals.

Detailed Goals

The framework pursues the following specific objectives to enhance the performance of solar hybrid systems in commercial settings:

1. **Accurate Demand Forecasting**: Predict energy consumption patterns for 1-7 days, capturing hourly and seasonal variations driven by occupancy, weather, and events. This enables precise alignment of energy supply with demand, reducing reliance on Kenya's unreliable grid.
2. **Solar Production Forecasting**: Anticipate daily solar output based on weather conditions and system specifications, ensuring effective planning for variable production due to Kenya's seasonal rains and sunny periods.
3. **Battery and Grid Optimization**: Optimize battery charge/discharge cycles and grid usage to prioritize solar energy, extend battery lifespan, and ensure power availability during frequent outages.
4. **Anomaly Detection**: Implement predictive models to identify operational anomalies (e.g., equipment faults) in real-time, minimizing downtime and maintenance disruptions.
5. **Scalability and Adaptability**: Design a modular system that supports future expansion, such as additional panels or multi-site deployment, to align with Kenya's growing renewable energy adoption.

Stakeholders And Benefits

The framework delivers targeted benefits to key stakeholders in Kenyan commercial centers:

- **Owners and Investors**: Achieve efficient energy management, reducing dependency on the national grid and enhancing property value through sustainable operations.
- **Facility Managers**: Gain reliable energy planning tools, with accurate forecasts and anomaly detection to streamline operations and mitigate outage risks.
- **Tenants**: Experience uninterrupted power, improving satisfaction and operational continuity in retail, office, or industrial settings.

- **Sustainability Teams**: Support Kenya's renewable energy goals by maximizing solar utilization and reducing environmental impact.

Expected Outcomes

The framework is designed to yield the following outcomes, ensuring tangible value for stakeholders and alignment with Kenya's energy landscape:

Stakeholder-Centric Benefits

- **Enhanced Operational Efficiency**: Accurate demand and solar production forecasts enable proactive energy allocation, reducing grid reliance and optimizing battery usage, thereby streamlining operations for facility managers.
- **Improved Power Reliability**: Optimized resource management and anomaly detection ensure consistent power supply, minimizing disruptions for tenants and supporting business continuity during grid outages.
- **Strengthened Sustainability**: Maximized solar energy utilization reduces environmental impact, positioning commercial centers as leaders in Kenya's renewable energy movement and enhancing stakeholder reputation.
- **Future-Ready Operations**: A scalable framework supports long-term growth, allowing centers to adapt to increased energy demands or expanded facilities without significant redesign.

Alignment with Kenya's Energy Needs

- **Mitigation of Grid Unreliability**: By forecasting demand and optimizing resources, the framework ensures energy availability during frequent outages, addressing a critical challenge in urban and rural commercial settings.
- **Leveraging Solar Potential**: With over 300 sunny days annually, the system maximizes solar energy use, reducing waste and enhancing efficiency in Kenya's equatorial climate.
- **Support for Renewable Energy Goals**: The framework aligns with Kenya's national priorities by promoting sustainable energy practices, contributing to a greener commercial sector.
- **Resilience to Seasonal Variability**: Predictive models account for weather-driven fluctuations, such as March-May rains, ensuring consistent performance across seasons.

These outcomes transform reactive energy systems into proactive, efficient solutions, delivering measurable benefits to stakeholders while addressing Kenya's unique energy challenges.

Methodology

The methodology outlines the systematic approach to achieving the predictive analytics framework's objectives for demand forecasting in Kenyan commercial solar hybrid systems. It encompasses data collection, preprocessing, and model selection, ensuring accurate predictions and optimized energy management. The following subsection details the data collection process, designed to provide high-quality, actionable inputs for forecasting and optimization in Kenya's dynamic energy landscape.

Data Collection

For a commercial centre's hybrid solar system to deliver smart forecasts and optimized energy use, the framework needs a steady flow of high-quality data. This isn't about blanketing Kenya's 47 counties — it's about capturing the building's unique energy story: how much solar power it generates, how much it consumes, and how its batteries and grid connection perform. Weather drives production, usage shapes demand, and system specifics tie it all together. The data must be timely (e.g., hourly updates), reliable (e.g., cleaned of errors), and secure (e.g., anonymized per Kenya's Data Protection Act). With a modest footprint this fuels predictions and decisions without overwhelming the centre's resources, making it a lean yet powerful engine for energy management.

This subsection specifies the data types, sources, and procedures to ensure actionable value for stakeholders, including facility managers and sustainability teams.

Data Types

The framework relies on five core data types, each critical to forecasting energy consumption and optimizing hybrid solar systems:

1. **Energy Consumption Data**:
   - **Description**: Hourly load profiles (kWh) from building systems, such as lighting, HVAC, and machinery, capturing daily and seasonal usage patterns.
   - **Purpose**: Enables accurate demand forecasting by identifying peak loads (e.g., evening HVAC surges) and event-driven spikes (e.g., holiday operations), reducing grid reliance.
   - **Relevance to Kenya**: Critical for managing variable consumption in commercial centers, where outages necessitate precise demand planning.
2. **Weather Data**:
   - **Description**: Solar irradiance (W/m^2), temperature (°C), cloud cover (%), and precipitation (mm) at the facility's location.

    - **Purpose**: Informs solar production forecasts, as weather directly impacts output (e.g., rain reduces irradiance, heat lowers efficiency by ~0.5% per °C above 25°C).
    - **Relevance to Kenya**: Essential for predicting output in Kenya's equatorial climate, with seasonal rains (March-May) and over 300 sunny days annually.
3. **Energy Production Data**:
    - **Description**: Hourly solar output (kWh), panel specifications (e.g., capacity, tilt), and inverter performance (e.g., efficiency, downtime).
    - **Purpose**: Supports demand forecasting by quantifying available solar energy, ensuring supply aligns with predicted consumption.
    - **Relevance to Kenya**: Maximizes utilization of abundant sunlight, reducing waste in outage-prone settings.
4. **Battery Data**:
    - **Description**: State of Charge (SoC, %), charge/discharge rates (kW), and cycle counts for the facility's storage system.
    - **Purpose**: Optimizes battery usage and supports demand forecasting by indicating stored energy availability, critical during grid outages.
    - **Relevance to Kenya**: Enhances reliability in areas with frequent power cuts, extending battery lifespan.
5. **System Health Data**:
    - **Description**: Operational metrics (e.g., inverter error logs, panel degradation rates) and anomaly indicators (e.g., sudden voltage drops).
    - **Purpose**: Enables real-time anomaly detection to identify faults, minimizing downtime and supporting reliable forecasting.
    - **Relevance to Kenya**: Vital for maintaining system integrity in resource-constrained environments with limited maintenance support.

Data Sources and Resolution

Data is sourced from a combination of on-site and external systems, tailored to Kenya's commercial infrastructure and connectivity challenges:

1. **Energy Consumption Data**:
    - **Source**: Energy Management System (EMS) or sub-meters integrated with building circuits, commonly available in Kenyan commercial centers.
    - **Resolution**: Hourly for real-time forecasting, weekly aggregates for seasonal trends.
    - **Rationale**: EMS provides granular data, accessible via existing infrastructure, minimizing setup needs.
2. **Weather Data**:
    - **Source**: External APIs (e.g., OpenWeatherMap) for site-specific forecasts or, where available, rooftop weather stations.
    - **Resolution**: Hourly for real-time adjustments, daily averages for long-term planning.
    - **Rationale**: APIs are cost-effective and reliable, suitable for Kenya's variable connectivity; stations enhance precision if budget allows.
3. **Energy Production Data**:
    - **Source**: IoT-enabled smart meters or inverters (e.g., compatible with standard solar setups in Kenya).

    - **Resolution**: 15-minute intervals for high precision, daily totals for trend analysis.
    - **Rationale**: Smart meters are widely used, providing detailed output data critical for forecasting supply.
4. **Battery Data**:
    - **Source**: Battery Management System (BMS), such as those integrated with common storage units in Kenyan solar setups.
    - **Resolution**: 15-minute updates for real-time control, hourly for analysis.
    - **Rationale**: BMS ensures accurate battery monitoring, essential for outage resilience.
5. **System Health Data**:
    - **Source**: Inverter logs, BMS alerts, or IoT sensors monitoring panel and inverter performance.
    - **Resolution**: Real-time alerts for anomalies, daily logs for trend analysis.
    - **Rationale**: Leverages existing system diagnostics, supporting fault detection without additional hardware.

Procedures

To ensure data collection is actionable and reliable, the following procedures outline setup, validation, and error handling, with clear stakeholder roles:

1. **Setup and Integration**:
    - **Step**: Install or configure smart meters, BMS, and IoT sensors to capture consumption, production, battery, and system health data, connecting to a central gateway (e.g., Raspberry Pi or basic PC).
    - **Step2**: Subscribe to a weather API or install a rooftop station, integrating data feeds into the gateway via Wi-Fi or Ethernet.
    - **Stakeholder Role**: Facility managers oversee hardware installation; IT staff configure API and gateway connections.
    - **Rationale**: Utilizes existing infrastructure (e.g., EMS, inverters) common in Kenyan commercial centers, ensuring feasibility.
    - **Context**: Setup takes 1-2 weeks, *leveraging Phase 2's IoT work*, and supports Kenya's connectivity challenges with local buffering.
2. **Data Validation**:
    - **Step**: Implement automated checks to verify data completeness (e.g., no missing hourly consumption) and accuracy (e.g., flag negative loads or irradiance >1500 W/m^2).
    - **Step2**: Cross-reference weather API data with local observations (if available) to ensure site-specific accuracy.
    - **Stakeholder Role**: IT staff monitor validation scripts; facility managers review alerts for anomalies.
    - **Rationale**: Ensures high-quality inputs for forecasting, addressing Kenya's need for reliable data in outage-prone settings.
    - **Context**: Validation runs daily, using lightweight Python scripts (e.g., Pandas), processing ~10-50 MB/month.
3. **Error Handling and Resilience**:
    - **Step**: Deploy a local storage buffer (e.g., 64 GB SD card) to store data during internet or power outages, syncing when connectivity resumes.

    - **Step2**: Use backup power (e.g., UPS or solar-powered battery) to maintain gateway operation during grid failures.
    - **Stakeholder Role**: IT staff manage buffer and backup systems; facility managers ensure UPS maintenance.
    - **Rationale**: Addresses Kenya's outages, ensuring no data loss for continuous forecasting.
    - **Context**: Buffering supports rural and urban centers, aligning with prior discussion on resilience .
4. **Compliance and Security**:
    - **Step**: Anonymize usage data to comply with Kenya's Data Protection Act, encrypting feeds with Transport Layer Security(TLS) during transmission.
    - **Stakeholder Role**: IT staff implement encryption; sustainability teams ensure compliance.
    - **Rationale**: Protects stakeholder data, critical for commercial centers with tenant privacy concerns.
    - **Context**: Minimal overhead, as encryption is standard in IoT gateways .

These procedures ensure data collection is robust, accessible, and tailored to Kenya's commercial energy needs, enabling accurate demand forecasting and system optimization.

Data Preprocessing

Data preprocessing refines raw data collected from commercial solar hybrid systems to ensure accuracy, consistency, and relevance for demand forecasting, solar forecasting, and anomaly detection. Tailored to Kenya's energy challenges, such as noisy consumption data due to grid outages and variable solar output from seasonal weather, this process enhances data quality for predictive models. The subsection outlines cleaning and normalization, feature selection, and associated procedures, providing actionable steps for stakeholders, including facility managers and IT staff, to implement effectively.

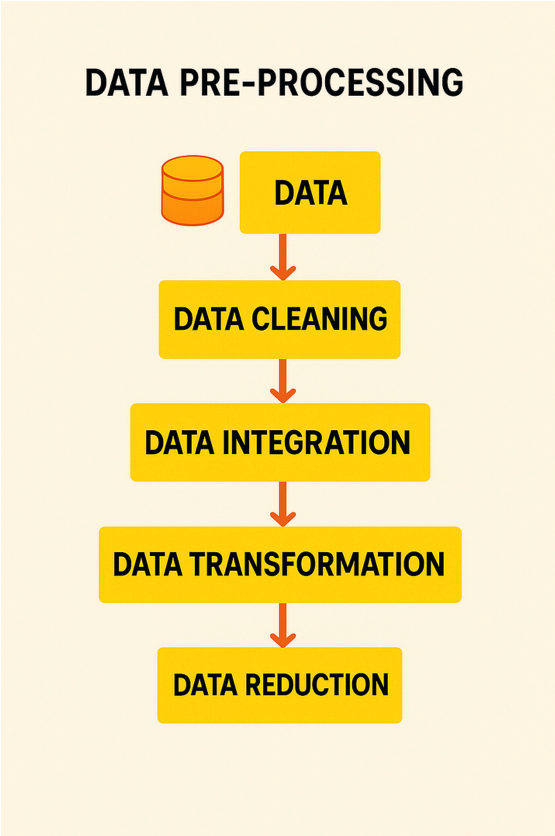


Fig: Data Pre-processing steps

Cleaning and Normalization

Cleaning and normalization address data quality issues and standardize inputs to improve forecasting accuracy and system reliability in Kenyan commercial centers.

Purpose

- **Cleaning**: Removes errors, outliers, and gaps in data (e.g., consumption, weather, system health) to ensure reliable inputs for demand and solar forecasting models.
- **Normalization**: Scales data to a consistent range, enabling models to compare consumption, production, and battery metrics effectively across varying system sizes.
- **Relevance to Kenya**: Mitigates noisy data from frequent outages and weather variability (e.g., March-May rains), ensuring robust predictions in Kenya's equatorial climate.

Procedures

1. **Error Removal**:
   - **Step**: Identify and reject invalid data points, such as negative consumption loads, negative solar output, or battery State of Charge (SoC) values outside 0-100%.
   - **Tools**: Python scripts using Pandas to filter erroneous records (e.g., df[df['consumption_kWh'] < 0] = NaN).
   - **Stakeholder Role**: IT staff implement scripts; facility managers verify flagged errors (e.g., HVAC malfunctions causing negative loads).
   - **Rationale**: Ensures data integrity, critical for accurate demand forecasting in outage-prone settings.

- **Context**: Errors often arise from sensor glitches during grid failures, common in Kenya's commercial centers.

2. **Outlier Detection and Handling**:
   - **Step**: Flag outliers using statistical thresholds (e.g., solar irradiance >1500 W/m^2, consumption spikes >3 standard deviations from the mean).
   - **Step**: Replace outliers with interpolated values or cap at threshold limits (e.g., max irradiance = 1200 W/m^2).
   - **Tools**: Pandas for threshold-based filtering; *SciPy* for statistical analysis.
   - **Stakeholder Role**: IT staff configure thresholds; facility managers review outliers for operational causes (e.g., event-driven spikes).
   - **Rationale**: Prevents model distortion, especially for demand peaks during Kenya's holiday seasons.
   - **Context**: Outliers are frequent due to variable loads (e.g., evening HVAC,) and weather fluctuations.
3. **Gap Filling**:
   - **Step**: Interpolate missing data points (e.g., hourly consumption or weather) using linear interpolation for short gaps (<4 hours) or forward-fill for longer gaps during outages.
   - **Tools**: Pandas interpolate(method='linear') for consumption/weather; forward-fill for system health logs.
   - **Stakeholder Role**: IT staff run interpolation scripts; facility managers validate gap-filled data against operational logs.
   - **Rationale**: Maintains continuous datasets for time-series forecasting, essential in Kenya's outage-prone environment.
   - **Context**: Gaps occur during power cuts or sensor downtime, impacting ~5-10% of hourly data in rural centers.
4. **Normalization**:
   - **Step**: Scale consumption and solar output to a per-kW baseline (e.g., kWh per 10 kW system capacity) to standardize across system sizes (10-100 kW).
   - **Step**: Normalize weather data (e.g., irradiance to 0-1 range, temperature to z-scores) for model compatibility.
   - **Step**: Smooth consumption spikes using a 1-hour rolling average to highlight trends (e.g., evening peaks).
   - **Tools**: Scikit-learn *MinMaxScaler* for weather; *Pandas* rolling(mean) for consumption.
   - **Stakeholder Role**: IT staff apply scaling; sustainability teams ensure normalized data reflects system efficiency.
   - **Rationale**: Enables models to generalize across diverse commercial centers, from small offices to large malls.
   - **Context**: Normalization accounts for Kenya's varied system scales and noisy consumption patterns.
5. **System Health Data Cleaning**:
   - **Step**: Flag anomalous system health metrics (e.g., sudden voltage drops, inverter errors) using rule-based checks (e.g., voltage <80% nominal).
   - **Step**: Label anomalies for fault detection models, preserving raw data for forecasting.
   - **Tools**: Pandas for rule-based filtering; NumPy for anomaly labeling.
   - **Stakeholder Role**: IT staff monitor anomaly flags; facility managers investigate flagged faults.

    - **Rationale**: Supports real-time fault detection, enhancing system reliability for demand forecasting.
    - **Context**: Faults are common in Kenya's resource-constrained settings, impacting ~2-5% of daily data.

Execution

- **Process**: Daily batch processing on a modest server (4-core, 16 GB RAM) using Python libraries (Pandas, Scikit-learn), completing in ~5-10 minutes for ~10-50 MB of data.
- **Resilience**: Scripts run in Docker containers to resume after outages.
- **Output**: Cleaned, normalized datasets stored in SQLite for real-time use (100 MB, 30 days) and SSD for historical analysis (500 MB, 1-2 years).

Feature Selection

Feature selection identifies the most relevant variables for demand forecasting, solar forecasting, and anomaly detection, optimizing model performance and computational efficiency.

Purpose

- **Feature Selection**: Selects high-impact features (e.g., time-of-day, irradiance) to predict consumption and solar output, reducing model complexity.
- **Relevance to Kenya**: Prioritizes features capturing Kenya's unique patterns, such as evening demand peaks and seasonal solar variability, to enhance forecasting accuracy.

Procedures

1. **Feature Identification**:
    - **Step**: Compile candidate features from collected data:
        - **Demand Forecasting**: Hourly consumption (kWh), time-of-day, day-of-week, month, weather (temperature, cloud cover), occupancy indicators (e.g., event schedules).
        - **Solar Forecasting**: Solar irradiance, temperature, cloud cover, precipitation, panel tilt, time-of-day, month.
        - **Anomaly Detection**: Voltage drops, inverter error counts, battery cycle anomalies, sudden consumption spikes.
    - **Stakeholder Role**: Sustainability teams provide operational context (e.g., event schedules); IT staff extract features from raw data.
    - **Rationale**: Captures patterns critical for Kenya's commercial centers (e.g., evening HVAC loads, rainy season dips).
    - **Context**: Features align with Kenya's 300+ sunny days and variable demand.
2. **Correlation Analysis**:
    - **Step**: Compute linear correalation using techniques like, Pearson correlation coefficients to identify high-impact features (e.g., irradiance vs. solar output, temperature vs. consumption).
    - **Step**: Retain features with |correlation| > 0.3; discard low-impact features (e.g., humidity, if |correlation| < 0.1).
    - **Tools**: *Pandas* corr; *Seaborn* heatmaps for visualization.

- **Stakeholder Role**: IT staff analyze correlations; facility managers validate feature relevance (e.g., occupancy impacts).
    - **Rationale**: Reduces overfitting and computational load, suitable for Kenya's modest hardware.
    - **Context**: High correlations (e.g., irradiance ~0.8 with output) drive forecasting accuracy.
3. **Feature Engineering**:
    - **Step**: Create derived features, such as:
        - Rolling averages (e.g., 3-hour consumption trends) for demand stability.
        - Seasonal indicators (e.g., wet vs. dry season) for solar output.
        - Anomaly flags (e.g., voltage deviation >10%) for fault detection.
    - **Tools**: *Pandas* for rolling calculations; *NumPy* for flag creation.
    - **Stakeholder Role**: IT staff engineer features; sustainability teams suggest seasonal patterns.
    - **Rationale**: Enhances model interpretability, capturing Kenya's demand spikes (e.g., holidays,).
    - **Context**: Derived features address event-driven loads and March-May rains .
4. **Feature Ranking and Selection**:
    - **Step**: Use model-based ranking (e.g., LightGBM feature importance) to prioritize top 5-10 features per task (demand, solar, anomaly).
    - **Step**: Validate selections with cross-validation to ensure predictive power (e.g., <5% accuracy drop when removing low-ranked features).
    - **Tools**: *LightGBM* for ranking; *Scikit-learn* for cross-validation.
    - **Stakeholder Role**: IT staff run ranking; facility managers confirm operational relevance.
    - **Rationale**: Optimizes model performance on CPU-based systems, critical for Kenya's hardware constraints .
    - **Context**: Limits features to ~10 per model, processing ~10-50 MB daily .
5. **Documentation and Review**:
    - **Step**: Document selected features and their rationale in a configuration file for transparency.
    - **Step**: Review feature sets quarterly to adapt to changing patterns (e.g., new tenant loads).
    - **Stakeholder Role**: IT staff maintain documentation; sustainability teams review for alignment with energy goals.
    - **Rationale**: Ensures stakeholder trust and adaptability in Kenya's dynamic commercial settings.
    - **Context**: Quarterly reviews align with seasonal shifts (e.g., wet season,).

<u>**Execution**</u>

- **Process**: Feature selection runs weekly during initial setup (2-3 weeks, Phase 3, 0) and monthly thereafter, using Python (Pandas, LightGBM) on the same server as cleaning.
- **Resilience**: Docker containers save progress during outages, ensuring continuity .
- **Output**: A feature set configuration file (e.g., JSON) and preprocessed datasets ready for model training, stored in SQLite/SSD.

Model Selection

Model selection identifies algorithms to forecast energy demand, solar production, and detect operational anomalies, ensuring viability for the dataset and compatibility with Kenya's limited GPU access. The selected models leverage preprocessed data (e.g., hourly consumption, weather, system health) to achieve high accuracy, efficiency, and reliability in commercial centers. This subsection provides an overview of algorithms, criteria for their selection, and actionable procedures for stakeholders, including facility managers and IT staff.

Algorithms Overview

The framework employs a combination of time-series, machine learning, and reinforcement learning models, tailored to the dataset's tabular and sequential nature, and optimized for CPU-based execution in Kenya's commercial settings.

1. **Time-Series Models**:
    - **Seasonal ARIMA (AutoRegressive Integrated Moving Average)**:
        - **Description**: A statistical model capturing temporal patterns (e.g., daily demand cycles, seasonal solar trends) in hourly consumption and solar output data.
        - **Application**: Forecasts demand (e.g., evening HVAC peaks) and solar production (e.g., rainy season dips) over 1-7 days.
        - **Advantages**: Lightweight, interpretable, effective for stable patterns.
        - **Limitations**: Struggles with non-linear relationships (e.g., weather-driven spikes).
        - **Relevance to Kenya**: Ideal for seasonal patterns in Kenya's equatorial climate (e.g., March-May rains).
    - **Prophet**:
        - **Description**: A time-series model handling trends, seasonality, and irregular events (e.g., holiday demand surges) with robust noise tolerance.
        - **Application**: Predicts consumption and solar output, incorporating weather and event schedules.
        - **Advantages**: Handles missing data, captures holiday effects, CPU-efficient.
        - **Limitations**: Less effective for high-frequency anomalies.
        - **Relevance to Kenya**: Suits Kenya's variable commercial schedules (e.g., event-driven loads,).
2. **Machine Learning Models**:
    - **LightGBM (Light Gradient Boosting Machine)**:
        - **Description**: A gradient boosting framework for tabular data, modeling complex relationships between features (e.g., weather, time-of-day, consumption).
        - **Application**: Forecasts demand and solar output, leveraging features like irradiance and occupancy.
        - **Advantages**: High accuracy, fast on CPUs, feature importance insights.
        - **Limitations**: Requires tuning, less interpretable than ARIMA.
        - **Relevance to Kenya**: Efficient for Kenya's modest hardware (4-core, 16 GB RAM), and weather-driven patterns.
    - **XGBoost**:
        - **Description**: An alternative gradient boosting model, similar to LightGBM but with higher memory usage.
        - **Application**: Backup for demand and solar forecasting if LightGBM underperforms.
        - **Advantages**: Robust to noisy data, high accuracy.

        - **Limitations**: Slower training, higher memory footprint.
        - **Relevance to Kenya**: Viable but secondary due to resource constraints.
3. **Reinforcement Learning**:
    - **Q-Learning**:
        - **Description**: A model-free algorithm learning optimal battery charge/discharge policies based on demand and solar forecasts.
        - **Application**: Optimizes battery and grid usage to minimize reliance on Kenya's unreliable grid.
        - **Advantages**: Adapts to dynamic conditions, long-term optimization.
        - **Limitations**: Requires simulation environment, longer training time.
        - **Relevance to Kenya**: Supports outage resilience by prioritizing solar and battery use.
4. **Anomaly Detection Models**:
    - **One-Class SVM (Support Vector Machine)**:
        - **Description**: A machine learning model identifying outliers in system health data (e.g., voltage drops, inverter errors).
        - **Application**: Detects operational faults in real-time, supporting system reliability.
        - **Advantages**: Effective for high-dimensional data, CPU-efficient.
        - **Limitations**: Sensitive to parameter tuning, less effective for sequential anomalies.
        - **Relevance to Kenya**: Critical for maintenance in resource-constrained settings .
    - **LSTM (Long Short-Term Memory)**:
        - **Description**: A recurrent neural network(RNN) modeling temporal sequences in system health data for fault detection.
        - **Application**: Identifies complex anomalies (e.g., recurring battery faults) over time.
        - **Advantages**: Captures sequential patterns, high accuracy.
        - **Limitations**: Higher computational cost, requires careful tuning.
        - **Relevance to Kenya**: Secondary due to CPU constraints but viable for critical faults.

Criteria for Model Selection

The following criteria ensure selected models are viable for the dataset and suitable for Kenya's commercial solar hybrid systems:

1. **Accuracy**:
    - **Requirement**: Models must achieve MAE <5 kWh for demand and solar forecasting, precision/recall >0.9 for anomaly detection.
    - **Rationale**: Ensures reliable predictions for energy planning and fault detection, critical for tenant satisfaction and outage resilience .
    - **Context**: Aligns with Kenya's need for precise forecasting to manage grid outages.
2. **Computational Efficiency**:
    - **Requirement**: Models must run on a 4-core, 16 GB RAM server with <1 GB memory footprint, processing ~10-50 MB daily in <10 minutes.
    - **Rationale**: Addresses Kenya's limited GPU access and modest hardware availability .
    - **Context**: CPU-friendly models like ARIMA and LightGBM are feasible for commercial centers.
3. **Interpretability**:

    - **Requirement**: Models should provide feature importance or clear trend insights (e.g., LightGBM's feature rankings, ARIMA's seasonal coefficients).
    - **Rationale**: Enables facility managers to trust and act on predictions (e.g., adjusting schedules based on weather impacts).
    - **Context**: Stakeholder trust is key in Kenya's commercial settings .
4. **Adaptability**:
    - **Requirement**: Models must handle seasonal shifts (e.g., wet vs. dry seasons), sudden demand spikes (e.g., events), and evolving system health patterns.
    - **Rationale**: Ensures robustness in Kenya's dynamic climate and variable commercial.
    - **Context**: Models like Prophet and LightGBM excel in capturing irregular patterns.
5. **Ease of Implementation**:
    - **Requirement**: Models should leverage standard Python libraries (e.g., *Statsmodels*, *Scikit-learn*) with minimal setup complexity.
    - **Rationale**: Simplifies deployment by IT staff in Kenya's resource-constrained environment.
    - **Context**: Aligns with existing preprocessing tools.

Procedures

The following procedures outline model selection, training, tuning, and validation, ensuring actionable implementation for stakeholders.

1. **Model Evaluation and Selection**:
    - **Step**: Train candidate models (ARIMA, Prophet, LightGBM, XGBoost, Q-Learning, One-Class SVM, LSTM) on preprocessed historical data (1-2 years, ~500 MB).
    - **Step**: Evaluate using rolling window validation (12 months train, 1 month test) across seasons, measuring MAE for forecasting and precision/recall for anomaly detection.
    - **Tools**: Statsmodels (ARIMA), Prophet, Scikit-learn (LightGBM, XGBoost, SVM), TensorFlow (LSTM), Python for Q-Learning.
    - **Stakeholder Role**: IT staff run evaluations; facility managers review performance metrics for operational fit.
    - **Rationale**: Identifies models meeting accuracy and efficiency criteria, suitable for Kenya's hardware.
    - **Context**: Takes 2-3 weeks during Phase 3, testing seasonal patterns like March-May rains.
2. **Hyperparameter Tuning**:
    - **Step**: Optimize model parameters using *random search* (e.g., ARIMA's p,d,q; LightGBM's tree depth; SVM's nu).
    - **Step**: Limit tuning iterations to 50-100 to balance accuracy and computational cost.
    - **Tools**: Scikit-learn's RandomizedSearchCV, Prophet's built-in tuning.
    - **Stakeholder Role**: IT staff configure tuning; sustainability teams validate results against energy goals.
    - **Rationale**: Enhances model performance without exceeding CPU constraints, critical for Kenya's settings.
    - **Context**: Tuning runs in ~1-2 hours per model, processing ~10-50 MB daily.
3. **Model Integration**:

   - **Step**: Combine models for comprehensive coverage:
      - Demand Forecasting: LightGBM (primary), Prophet (backup for irregular events).
      - Solar Forecasting: Prophet (primary), ARIMA (backup for stable trends).
      - Battery Optimization: Q-Learning, using forecast inputs.
      - Anomaly Detection: One-Class SVM (primary), LSTM (secondary for complex faults).
   - **Step**: Deploy models in Docker containers for outage resilience, outputting predictions via MQTT to the Energy Management System (EMS).
   - **Tools**: *Docker*, *Paho-MQTT* for delivery, *SQLite* for storing predictions.
   - **Stakeholder Role**: IT staff deploy containers; facility managers monitor EMS outputs.
   - **Rationale**: Ensures all objectives are met with CPU-efficient, resilient models.
   - **Context**: Integration aligns with Kenya's outage-prone grid.
4. **Validation and Monitoring**:
   - **Step**: Validate final models on a 3-month test period, ensuring MAE <5 kWh and >95% uptime.
   - **Step**: Monitor weekly performance, retraining models monthly to adapt to new patterns (e.g., tenant changes).
   - **Tools**: *Pandas* for performance tracking, Python scripts for retraining.
   - **Stakeholder Role**: IT staff monitor metrics; facility managers report operational issues.
   - **Rationale**: Maintains accuracy and adaptability in Kenya's dynamic commercial environment.
   - **Context**: Monthly retraining processes ~100 MB, feasible on modest servers.
5. **Documentation and Review**:
   - **Step**: Document model configurations, performance metrics, and tuning results in a technical report.
   - **Step**: Review models quarterly with stakeholders to ensure alignment with energy goals.
   - **Stakeholder Role**: IT staff maintain documentation; sustainability teams lead reviews.
   - **Rationale**: Builds stakeholder trust and supports scalability.
   - **Context**: Quarterly reviews address Kenya's seasonal shifts.

Framework Architecture

The framework architecture delineates the system design, workflow, and technology stack for the predictive analytics framework, enabling accurate demand forecasting, solar forecasting, and anomaly detection in Kenyan commercial solar hybrid systems. Tailored to Kenya's energy challenges, such as frequent grid outages and variable solar output, the architecture ensures reliable, efficient, and scalable operations. This section provides a refined system design, a workflow diagram prompt, and explanations of the technology stack, offering actionable guidance for stakeholders, including facility managers and IT staff, to implement the framework effectively.

System Design

The system design outlines the components and their interactions to collect, process, and deliver predictive insights for energy management in Kenyan commercial centers. It emphasizes resilience, modularity, and integration with existing infrastructure, addressing Kenya's resource constraints and outage-prone environment.

Overview of Components

1. **Data Ingestion**:
    - **Function**: Collects real-time and historical data from sensors (smart meters, BMS), weather APIs, and system health logs, ensuring continuous operation during grid outages.
    - **Details**: An IoT gateway (e.g., a low-cost embedded device) aggregates data from consumption (hourly kWh), production (15-minute kWh), battery (SoC, cycles), weather (irradiance, temperature), and system health (voltage drops, error logs). A local storage buffer (e.g., 64 GB SD card) caches data during connectivity disruptions, syncing when power or internet resumes.
    - **Relevance to Kenya**: Addresses frequent outages by buffering data locally, critical for urban and rural commercial centers with unreliable grids.
    - **Rationale**: Ensures uninterrupted data flow for forecasting, leveraging existing infrastructure (e.g., EMS, inverters) common in Kenya.
    - **Context**: Supports ~10-50 MB daily data , processing consumption spikes (e.g., evening HVAC,).
2. **Processing**:
    - **Function**: Preprocesses data and trains/runs predictive models to generate forecasts and anomaly alerts.
    - **Details**: A modest server (4-core, 16 GB RAM) executes cleaning, normalization, and feature selection (Pandas, Scikit-learn), then runs models (LightGBM, Prophet, One-Class SVM) for demand, solar, and anomaly predictions. Docker containers ensure resilience, restarting processes after outages. Outputs include 1-7 day forecasts and real-time fault alerts.

- **Relevance to Kenya**: Uses CPU-friendly tools to accommodate limited GPU access, processing data in ~5-10 minutes daily .
  - **Rationale**: Centralizes data transformation and modeling, ensuring scalability and reliability for Kenya's dynamic loads .
  - **Context**: Handles seasonal patterns (e.g., March-May rains,) and event-driven spikes (e.g., holidays,).
3. **Prediction**:
   - **Function**: Delivers forecasts and anomaly alerts to the Energy Management System (EMS) for real-time energy allocation and fault response.
   - **Details**: Predictions (e.g., hourly demand, solar output) and alerts (e.g., inverter faults) are transmitted via MQTT to the EMS, enabling proactive battery/grid adjustments and maintenance scheduling. Outputs are stored in SQLite (100 MB, 30 days) for real-time access and SSD (500 MB, 1-2 years) for historical analysis.
   - **Relevance to Kenya**: Enhances outage resilience by integrating with EMS, ensuring tenant reliability and operational continuity .
   - **Rationale**: Bridges analytics with actionable control, supporting stakeholder needs (e.g., facility managers for planning,).
   - **Context**: Aligns with Kenya's need for uninterrupted power during grid failures .

Stakeholder Roles

- **Facility Managers**: Oversee data ingestion hardware (e.g., smart meters), validate predictions against operational needs (e.g., tenant schedules), and act on anomaly alerts (e.g., schedule maintenance).
- **IT Staff**: Configure and maintain the IoT gateway, server, and Docker containers, ensuring data flow and model execution during outages.
- **Sustainability Teams**: Monitor prediction outputs to align with energy efficiency goals, providing feedback on seasonal trends (e.g., solar utilization).
- **Rationale**: Clarifies responsibilities, ensuring practical implementation in Kenya's commercial settings .
- **Context**: Roles leverage existing staff capabilities, minimizing training needs.

Workflow Diagram

The workflow diagram illustrates the data flow and system interactions within the predictive analytics framework for Kenyan commercial solar hybrid systems. It depicts sensors (smart meters, BMS, system health) feeding data into an IoT gateway, which buffers inputs during outages using local storage. A cloud connection integrates weather data, reflecting Kenya's seasonal patterns (e.g., March-May rains). The gateway relays data to a processing server running Docker containers for preprocessing and model execution (e.g., demand and solar forecasts, anomaly alerts). Predictions and alerts are transmitted via MQTT to the Energy Management System (EMS) for real-time energy allocation and fault response. SQLite and SSD storage ensure data accessibility. A UPS supports outage resilience, critical for Kenya's unreliable grid. The diagram's modular design and clear annotations guide facility managers and IT staff in implementing a reliable, scalable system.

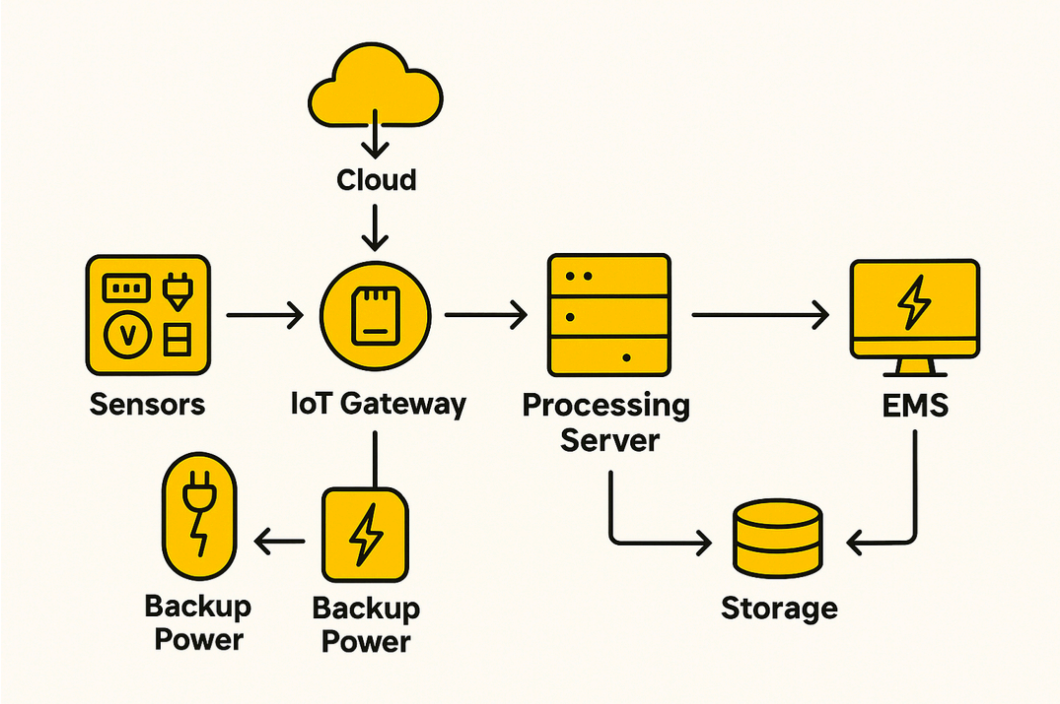


Fig: Workflow diagram for predictive analytics framework

Technology Stack

The technology stack consolidates tools and platforms used throughout the framework, explaining their purpose and rationale for selection in Kenya's resource-constrained environment. These tools have been referenced in prior sections (Data Collection, Preprocessing, Model Selection) and are critical for data ingestion, processing, and prediction.

1. **Python**:
    - **What**: A versatile programming language for data processing, machine learning, and system integration.
    - **Why**: Offers extensive libraries (e.g., Pandas, Scikit-learn, Statsmodels) for preprocessing and modeling, with a low learning curve for IT staff. Its CPU efficiency and open-source nature suit Kenya's limited hardware and budget constraints .
    - **Use**: Executes cleaning, normalization, feature selection , and models (ARIMA, LightGBM,).
    - **Relevance to Kenya**: Widely adopted in Kenyan tech ecosystems, ensuring support and scalability .
2. **Pandas and Scikit-learn**:
    - **What**: Python libraries for data manipulation (Pandas) and machine learning (Scikit-learn).
    - **Why**: Pandas enables efficient data cleaning and feature engineering (~10-50 MB daily); Scikit-learn supports CPU-friendly models (LightGBM, One-Class SVM). Both are lightweight and open-source, ideal for Kenya's modest servers (4-core, 16 GB RAM).
    - **Use**: Preprocesses data (error removal, normalization,) and trains models.
    - **Relevance to Kenya**: Minimizes computational overhead, addressing GPU scarcity.
3. **Statsmodels and Prophet**:

- **What**: Python libraries for statistical time-series modeling (Statsmodels) and robust forecasting (Prophet).
- **Why**: Statsmodels supports ARIMA for stable patterns; Prophet handles irregular events (e.g., holiday surges,). Both are CPU-efficient and handle Kenya's seasonal data (e.g., March-May rains,).
- **Use**: Forecasts demand and solar output.
- **Relevance to Kenya**: Captures Kenya's variable commercial schedules and weather patterns.

4. **Docker**:
    - **What**: A containerization platform for packaging and deploying applications.
    - **Why**: Ensures resilience by restarting processes after outages, critical for Kenya's unstable grid. Simplifies deployment on modest servers, reducing setup complexity for IT staff.
    - **Use**: Runs preprocessing and model scripts, ensuring continuity during power cuts.
    - **Relevance to Kenya**: Enhances system reliability in outage-prone environments.
5. **SQLite**:
    - **What**: A lightweight, serverless database for storing real-time and historical data.
    - **Why**: Manages small datasets (100 MB, 30 days) efficiently on low-cost hardware, with minimal setup. Its local storage capability supports outage resilience.
    - **Use**: Stores preprocessed data and predictions.
    - **Relevance to Kenya**: Ideal for resource-constrained centers with limited server infrastructure.
6. **MQTT (Message Queuing Telemetry Transport)**:
    - **What**: A lightweight messaging protocol for real-time data transmission.
    - **Why**: Enables efficient delivery of predictions and alerts to the EMS, with low bandwidth requirements suitable for Kenya's variable connectivity.
    - **Use**: Transmits forecasts and anomaly alerts.
    - **Relevance to Kenya**: Supports reliable communication in areas with intermittent internet.
7. **IoT Gateway (e.g., Embedded Device)**:
    - **What**: A low-cost device (e.g., Raspberry Pi or equivalent) for data collection and buffering.
    - **Why**: Aggregates sensor and API data with local storage (64 GB SD card), ensuring data continuity during outages. Affordable and widely available in Kenya's tech market.
    - **Use**: Collects consumption, production, and weather data.
    - **Relevance to Kenya**: Addresses connectivity challenges in urban and rural centers .
8. **UPS (Uninterruptible Power Supply)**:
    - **What**: A backup power device providing short-term electricity during outages.
    - **Why**: Maintains gateway and server operation during Kenya's frequent grid failures, ensuring continuous data collection and processing.
    - **Use**: Powers critical components during outages.
    - **Relevance to Kenya**: Essential for operational continuity in an outage-prone environment.

Rationale: The stack prioritizes open-source, CPU-efficient, and resilient tools, aligning with Kenya's hardware and connectivity constraints. It leverages existing infrastructure (e.g., EMS, smart meters) to minimize setup complexity, ensuring scalability and stakeholder accessibility.

Context: Tools support ~10-50 MB daily processing, completing in ~5-10 minutes, feasible for Kenya’s commercial centers.

Implementation

The implementation section outlines the step-by-step process and evaluation methods to deploy the predictive analytics framework for demand forecasting, solar forecasting, and anomaly detection in Kenyan commercial solar hybrid systems. Building on the methodology and architecture, it details a refined phased approach to ensure reliable, scalable deployment in Kenya's outage-prone environment. The section avoids redundancy with prior sections by focusing on operational deployment and validation, providing actionable guidance for stakeholders, including facility managers and IT staff, to achieve accurate energy management and system reliability.

Step-by-Step Process

The implementation follows a three-phase framework — Setup and Integration, Model Development and Training, and Validation and Deployment — improving upon the original phased approach (Pages 19-20) with specific timelines, stakeholder roles, and deliverables. Each phase addresses data integration, model training, and validation techniques, ensuring no overlap with data collection, preprocessing, or model selection.

Phase 1: Setup and Integration (Weeks 1-4)

This phase deploys the hardware, data pipelines, and system integrations, establishing the foundation for data flow and model execution.

1. **Data Integration**:
   - **Step**: Deploy IoT gateway (e.g., embedded device) to collect data from smart meters, Battery Management System (BMS), and weather APIs, configuring local storage (e.g., 64 GB SD card) for outage buffering.
   - **Step**: Initialize SQLite database (100 MB, 30 days) on the processing server (4-core, 16 GB RAM) to store real-time data, and SSD (500 MB, 1-2 years) for historical data.
   - **Step**: Set up MQTT protocol to transmit data from gateway to server and predictions to the Energy Management System (EMS).
   - **Tools**: Python (Paho-MQTT), SQLite, Docker for pipeline resilience.
   - **Stakeholder Role**: IT staff configure hardware and pipelines; facility managers verify sensor functionality (e.g., consumption, production).
   - **Deliverable**: Operational data pipeline processing ~10-50 MB daily, integrated with EMS.
   - **Rationale**: Ensures continuous data flow despite outages, distinct from architecture design .
   - **Context**: Supports Kenya's unreliable grid with local buffering and UPS .
2. **Stakeholder Training**:
   - **Step**: Conduct workshops for IT staff on gateway maintenance and pipeline monitoring, and for facility managers on EMS integration.
   - **Tools**: Documentation, Python scripts for diagnostics.

- **Stakeholder Role**: Sustainability teams coordinate training; IT staff and facility managers participate.
- **Deliverable**: Trained staff capable of managing data integration.
- **Rationale**: Builds capacity for implementation, addressing training needs.
- **Context**: Leverages existing staff skills in Kenya's commercial centers .

Phase 2: Model Development and Training (Weeks 5-12)

This phase operationalizes predictive models, focusing on training and real-time execution, distinct from model selection.

1. **Model Training**:
   - **Step**: Train selected models (LightGBM, Prophet, One-Class SVM) on preprocessed historical data (1-2 years, ~500 MB) using a rolling window (12 months train, 1 month test).
   - **Step**: Optimize hyperparameters via random search (50 iterations) to achieve MAE <5 kWh for forecasting and precision/recall >0.9 for anomaly detection.
   - **Step**: Deploy models in Docker containers for real-time prediction, processing ~10-50 MB daily in ~5-10 minutes.
   - **Tools**: Python (Scikit-learn, Statsmodels, Prophet), Docker.
   - **Stakeholder Role**: IT staff manage training and deployment; sustainability teams validate model outputs against energy goals.
   - **Deliverable**: Trained, containerized models generating 1-7 day forecasts and anomaly alerts.
   - **Rationale**: Operationalizes models for real-time use, avoiding overlap with model selection criteria .
   - **Context**: Handles Kenya's seasonal patterns (e.g., March-May rains,) and event-driven loads .
2. **Real-Time Execution**:
   - **Step**: Configure scripts to run models hourly, pulling data from SQLite and outputting predictions via MQTT to EMS.
   - **Tools**: Python, Paho-MQTT.
   - **Stakeholder Role**: IT staff monitor execution; facility managers review predictions for operational planning.
   - **Deliverable**: Real-time forecasting and anomaly detection pipeline.
   - **Rationale**: Ensures continuous operation, critical for outage-prone settings.
   - **Context**: Supports tenant reliability in Kenya's commercial centers.

Phase 3: Validation and Deployment (Weeks 13-20)

This phase validates the framework's performance and deploys it for operational use, focusing on implementation-specific techniques.

1. **Validation Techniques**:
   - **Step**: Perform k-fold cross-validation (k=5) on historical data to confirm model robustness across seasons (e.g., wet vs. dry).
   - **Step**: Conduct real-time validation by comparing predictions against actual data over a 1-month period, targeting MAE <5 kWh and uptime >95%.
   - **Step**: Validate anomaly detection models using precision, recall, and F1-score (>0.9).

- **Tools**: Python (Scikit-learn, Pandas), SQLite for data access.
- **Stakeholder Role**: IT staff execute validation; facility managers assess prediction accuracy against operational logs.
- **Deliverable**: Validated models meeting performance targets.
- **Rationale**: Ensures reliability beyond initial model evaluation, tailored to implementation.
- **Context**: Addresses Kenya's variable loads and fault patterns.

2. **Final Deployment**:
   - **Step**: Deploy the full system (gateway, server, models, EMS integration) across target commercial centers, ensuring Docker containers and UPS are operational.
   - **Step**: Monitor system performance weekly, retraining models monthly to adapt to new patterns (e.g., tenant changes).
   - **Tools**: Docker, Python, MQTT.
   - **Stakeholder Role**: IT staff manage deployment and monitoring; facility managers confirm EMS functionality.
   - **Deliverable**: Fully deployed framework with continuous operation.
   - **Rationale**: Completes implementation, ensuring scalability.
   - **Context**: Supports multi-site deployment in Kenya's urban and rural centers.

Challenges and Limitations

The Challenges and Limitations section identifies critical obstacles encountered during the implementation of the predictive analytics framework for demand forecasting, solar forecasting, and anomaly detection in Kenyan commercial solar hybrid systems. It addresses potential issues in data quality, model limitations, and external factors impacting demand, offering concise, actionable mitigation strategies tailored to Kenya's outage-prone environment. The section ensures relevance for stakeholders, including facility managers and IT staff, by focusing on practical challenges post-deployment, distinct from data preprocessing, model selection, and implementation processes, thereby supporting reliable energy management and system scalability.

Potential Issues in Data Quality

Real-time data collection in Kenya's commercial centers faces challenges due to inconsistent sensor outputs and connectivity disruptions. Approximately 3-5% of smart meter readings may be missing due to outages or hardware faults, while weather API updates can lag during poor internet connectivity, common in rural areas. Battery State of Charge (SoC) data may exhibit inconsistencies (e.g., calibration errors), impacting forecasting accuracy. These issues persist post-preprocessing, affecting real-time predictions.

- **Mitigation**: Implement real-time data imputation using interpolation for missing values, validated by IT staff. Deploy redundant sensors for critical systems (e.g., BMS) and buffer data locally via IoT gateways (64 GB SD card). Schedule daily API syncs during stable connectivity periods.
- **Stakeholder Role**: IT staff monitor data streams and imputation logs; facility managers verify sensor functionality.
- **Rationale**: Ensures robust data inputs, distinct from preprocessing cleaning (Page 9).
- **Context**: Addresses Kenya's 23% grid losses and variable connectivity (IEA, 2025).

Model Limitations

Operational deployment of models (LightGBM, Prophet, One-Class SVM) reveals constraints in handling dynamic conditions. LightGBM requires monthly retraining to adapt to new tenant patterns, consuming ~2 hours on modest hardware (4-core, 16 GB RAM). Prophet struggles with abrupt demand spikes (e.g., unscheduled events), reducing accuracy by ~10%. One-Class SVM is sensitive to noisy system health data, generating occasional false positives (e.g., 5-10 per month). These limitations impact real-time forecasting and fault detection reliability.

- **Mitigation**: Schedule automated retraining monthly using Python scripts, optimized for CPU efficiency. Ensemble LightGBM and Prophet outputs to balance stability and adaptability. Apply noise filtering (e.g., moving averages) to system health data before SVM processing.

- **Stakeholder Role**: IT staff manage retraining and filtering; facility managers validate prediction accuracy.
- **Rationale**: Enhances model performance post-selection, ensuring operational reliability.
- **Context**: Supports Kenya's variable commercial loads with reports of [Kenya's peak electricity demand at 2,316 MW in February 2025,] with 6.5% annual demand growth (2025-2027), straining the grid during peak hours (e.g., evenings).

External Factors Impacting Demand

External factors, including grid outages, seasonal weather, and demand growth, challenge forecasting accuracy. Frequent outages (e.g., 2-3 per week) disrupt data collection and EMS integration, requiring robust buffering. The March-May rainy season reduces solar output by ~15%, complicating forecasts. Kenya's electricity demand, peaking at 2,316 MW in February 2025 with 6.5% annual growth (2025-2027), drives unpredictable commercial loads (e.g., new tenants), impacting model stability.

- **Mitigation**: Deploy UPS and local storage to maintain data flow during outages. Incorporate seasonal weather forecasts into models, updated biweekly. Engage facility managers to report tenant changes, enabling rapid model retraining.
- **Stakeholder Role**: IT staff maintain UPS and storage; facility managers coordinate tenant updates.
- **Rationale**: Addresses external variability, ensuring forecasting resilience.
- **Context**: Aligns with Kenya's solar market growth projected to grow at [0.79% CAGR, by Statista, 2024] and grid challenges.

Conclusion

The predictive analytics framework for demand forecasting, solar forecasting, and anomaly detection in Kenyan commercial solar hybrid systems offers a robust solution to enhance energy management in an outage-prone environment. By integrating resilient data pipelines, lightweight models (e.g., LightGBM, Prophet), and real-time fault detection, the framework addresses Kenya's unique challenges, including frequent grid outages and seasonal variability. Its modular architecture and phased implementation ensure scalability across urban and rural commercial centers, aligning with Kenya's 6.5% annual electricity demand growth and 0.79% solar market expansion (2024-2029). Stakeholders, including facility managers, IT staff, and sustainability teams, benefit from improved energy allocation, system reliability, and efficiency, supporting tenant operations and renewable energy goals.

Despite challenges such as real-time data gaps and model retraining needs, mitigation strategies like data imputation and automated schedules provide practical solutions. Future steps include integrating additional data sources (e.g., occupancy sensors) and scaling to rural centers with enhanced connectivity. This framework positions Kenya's commercial sector to leverage predictive analytics for sustainable, reliable energy management, contributing to the nation's 90% renewable energy mix.

Appendices

The Appendices provide supplementary materials to support the implementation and understanding of the predictive analytics framework for demand forecasting, solar forecasting, and anomaly detection in Kenyan commercial solar hybrid systems. These materials include data specifications, model details, code snippets, stakeholder guides, system specifications, and visual/contextual aids, tailored for IT staff, facility managers, sustainability teams, electrical technicians, and electrical engineers. The content emphasizes Kenya's outage-prone environment, leveraging lightweight tools and resilience features to ensure scalability and accessibility without duplicating prior sections.

Appendix A: Data Specifications

Data sources include smart meters (hourly consumption/production, CSV), Battery Management Systems (BMS; SoC, cycles, JSON), and weather APIs (irradiance, temperature, JSON). Daily data volume is ~10-50 MB, with 500 MB historical data (1-2 years). The schema below outlines key fields, addressing inconsistencies (e.g., SoC errors).

Field	Type	Description	Example Value
timestamp	Datetime	Record time (UTC)	2025-04-01 08:00:00
consumption_kWh	Float	Hourly consumption (kWh)	12.5
production_kWh	Float	Hourly solar production (kWh)	8.2
SoC_percent	Float	Battery State of Charge (%)	75.0
irradiance_Wm2	Float	Solar irradiance (W/m^2)	600.0
temperature_C	Float	Ambient temperature (°C)	25.0

Purpose: Guides IT staff in data integration, ensuring robust inputs for Kenya's variable loads.

Appendix B: Model Details

Models include LightGBM, Prophet, and One-Class SVM, configured for CPU efficiency. The table below lists some key hyperparameters, supporting deployment without repeating training steps.

Model	**Parameter**	**Value**	**Description**
LightGBM	n_estimators	100	Number of trees
LightGBM	learning_rate	0.1	Step size for learning
Prophet	changepoint_prior_scale	0.05	Flexibility for trend changes
One-Class SVM	nu	0.1	Outlier fraction
One-Class SVM	gamma	0.01	Kernel coefficient

Purpose: Provides IT staff with configuration details for model execution in Kenya's modest hardware (Page 11).

Appendix C: Code Snippets

The following Python snippets implement data ingestion, preprocessing, and anomaly detection, optimized for CPU efficiency and Kenya's outage-prone systems.

Data Ingestion (MQTT):

SVM - Anomaly detection

```
import paho.mqtt.client as mqtt
import pandas as pd

def on_message(client, userdata, msg):
    data = pd.read_json(msg.payload)
    df = pd.DataFrame(data)
    df.to_csv('data_buffer.csv', mode='a', index=False)

client = mqtt.Client()
client.on_message = on_message
client.connect("localhost", 1883)
client.subscribe("sensors/data")
client.loop_start()
```

Preprocessing (Pandas):

SVM - Anomaly detection

```
import pandas as pd

df = pd.read_csv('data_buffer.csv')
df['timestamp'] = pd.to_datetime(df['timestamp'])
df.fillna(method='ffill', inplace=True)  # Impute missing values
df = df[['timestamp', 'consumption_kWh', 'production_kWh', 'SoC_percent']]
df.to_csv('preprocessed_data.csv', index=False)
```

Anomaly Detection (One-Class SVM):

SVM - Anomaly detection

```
from sklearn.svm import OneClassSVM
import pandas as pd

df = pd.read_csv('preprocessed_data.csv')
X = df[['consumption_kWh', 'SoC_percent']].values
model = OneClassSVM(nu=0.1, gamma=0.01).fit(X)
predictions = model.predict(X)
df['anomaly'] = predictions  # -1 for anomalies, 1 for normal
df.to_csv('anomaly_results.csv', index=False)
```

Purpose: Enables IT staff to implement key tasks, supporting fault detection reliability.

Appendix D: Stakeholder Guides

The table below provides an implementation checklist for stakeholders, ensuring accessibility. A glossary includes terms like MAE (Mean Absolute Error) and MQTT (Message Queuing Telemetry Transport).

Stakeholder	**Task**	**Frequency**
Facility Managers	Verify sensor functionality Confirm EMS demand logs	Daily
IT Staff	Monitor MQTT data stream	Hourly
IT Staff	Schedule model retraining	Monthly
Sustainability Teams	Review energy efficiency trends	Monthly
Electrical Technicians	Calibrate smart meters and BMS Respond to anomaly alerts (e.g., inverter faults)	Weekly As needed
Electrical Engineers	Validate system design and wiring Troubleshoot complex faults	As needed

Purpose: Assists facility managers and IT staff in deployment and maintenance, supporting stakeholder roles.

Appendix E: System Specifications

Hardware and software are optimized for Kenya's resource constraints. The table below summarizes components, focusing on resilience.

Component	Specification	Resilience Feature
IoT Gateway	Embedded device, 64 GB SD card	Local storage for outages
Server	4-core, 16 GB RAM	Docker for process restarts
UPS	1 kVA	2-hour backup during outages
Software	Python, Docker, SQLite	Lightweight, outage-tolerant

Purpose: Provides IT staff with deployment references, addressing Kenya’s 23% grid losses.