Jurnal Mandiri IT
Vol. 14 No. 3, January (2026), pp. 294-307

ISSN 2301-8984 (Print), 2809-1884 (Online)
Published by Institute of Computer Science

Energy consumption prediction and optimization for Ki Hajar Dewantara student dormitory Using Extreme Gradient Boosting (XGBoost) algorithm

Jeremia Paskah Putra Sinaga[1], Jonson Manurung[2], M. Azhar Prabukusumo[3]
[1,2,3] Informatics, Faculty of Defense Technology and Engineering, Indonesia Defense University, Bogor, Indonesia

ARTICLE INFO

Article history:

Received Dec 14, 2025
Revised Dec 27, 2025
Accepted Jan 9, 2026

Keywords:

Building Energy Management;
Energy Optimization;
Machine Learning;
Predictive Analytics;
XGBoost.

ABSTRACT

Energy consumption optimization in student dormitories requires accurate prediction and strategic intervention strategies. This study presents a comprehensive prediction and optimization system for energy usage at Ki Hajar Dewantara Student Dormitory, Indonesia Defense University, utilizing Extreme Gradient Boosting (XGBoost) algorithm integrated with temporal operational scheduling features a novel approach for institutional dormitory energy forecasting. The system analyzes over 3,900 electrical devices across three dormitory buildings, incorporating temporal features and operational schedules to predict hourly energy consumption. The XGBoost model demonstrates excellent prediction performance with $R^2 = 0.9482$ and MAPE = 10.24%, significantly exceeding established benchmarks for building energy forecasting. Feature importance analysis reveals working hours as the dominant factor (>85%) influencing consumption patterns, followed by occupancy rate and temperature. The analysis identifies air conditioning systems as the primary energy consumer, accounting for over 80% of total consumption. The optimization framework identifies potential energy savings of approximately 28% through strategic device replacement and schedule modifications, translating to annual cost savings of over Rp 600 million with economically viable return on investment periods. This machine learning-based approach demonstrates practical applicability for student dormitory energy management and provides a replicable methodology adaptable to diverse residential institutional buildings in tropical climates.

Corresponding Author:

Jeremia Paskah Putra Sinaga,
Informatics,
Indonesia Defense University,
Kawasan IPSC Sentul, Sukahati, Kec. Citeureup, Kabupaten Bogor, Jawa Barat 16810, Indonesia.
Email: jeremiasinaga102@gmail.com

1. INTRODUCTION

Energy efficiency in student dormitories is critical as institutions face escalating costs and sustainability mandates. Dormitory buildings account for 20-30% of institutional energy consumption (Cui et al., 2024; Deshko et al., 2020). Traditional energy audits fail to capture dynamic patterns (Alizadegan et al., 2025). Machine learning algorithms enable continuous monitoring and data driven optimization (Baqer & Rashidi Khazaee, 2025; Yeboah Ofori et al., 2021). XGBoost demonstrates superior performance in energy prediction through non linear modeling and feature importance analysis (Dai & Huang, 2025; Wu et al., 2024).Among various machine learning algorithms, XGBoost is selected for this study due to its inherent advantages: gradient boosting architecture that handles non linear energy consumption patterns effectively, built in regularization preventing overfitting in temporal datasets, native support for feature importance ranking enabling transparent decision making for facility managers, and computational efficiency suitable for real time institutional deployment. These characteristics make XGBoost particularly appropriate for dormitory energy systems where interpretability and operational feasibility are equally important as prediction accuracy.

Recent studies applied machine learning to building energy management. Wu et al. (2024) achieved $R^2 > 0.90$ combining XGBoost with NSGA II. Dai & Huang (2025) demonstrated XGBoost superiority with MAPE < 12%. Makrygiorgou et al. (2020) applied deep learning for cooling load prediction. Lai et al. (2022) validated ensemble methods for residential forecasting. Roy et al. (2020)reviewed ML techniques achieving 5-15% MAPE. However, these studies focused on residential/commercial buildings, not institutional dormitories (Cui et al., 2024). Limited research integrated prediction with device level ROI recommendations (Saragih & Novimariono, 2020; Yesilyurt et al., 2024).

Despite the extensive literature on building energy prediction, a critical research gap exists at the intersection of institutional dormitory management, device level granularity, and economic optimization. This study addresses three fundamental novelties distinguishing it from prior work: First, the application context focuses specifically on educational dormitory environments with unique occupancy patterns, operational schedules, and multi building coordination requirements absent in typical residential or commercial settings. Second, the analytical granularity operates at individual device level across 32 equipment categories rather than aggregate building level consumption, enabling targeted replacement strategies and schedule modifications. Third, the methodological integration combines machine learning prediction with return on investment calculations, providing actionable economic guidance beyond statistical forecasting alone. These distinctions position this research as a pioneering framework for institutional energy management in developing country contexts where resource constraints demand evidence based prioritization.

Literature gaps exist in dormitory energy systems combining device level analysis with machine learning (Blanco et al., 2020; Walsh et al., 2020). Smart building optimization requires innovative solutions (Hanif et al., 2019). IoT enabled monitoring systems enhance real time energy management (Cabeza & Chàfer, 2020).This research develops a comprehensive XGBoost based system for student dormitories with specific objectives: (1) achieve $R^2 \geq 0.85$ and MAPE ≤ 15% prediction accuracy, (2) identify device level consumption patterns through feature importance analysis, (3) generate ROI based optimization recommendations, and (4) validate using 32 device types (3,900 units) from Ki Hajar Dewantara Dormitory.

This study integrates device inventory, temporal features, ML prediction, and economic optimization (Zhang et al., 2020;Gernaat et al., 2020). Key contributions: (1) reproducible methodology with 32 device types, 17 features, (2) $R^2 = 0.9482$, MAPE = 10.24% exceeding benchmarks, (3) 27.63% savings potential (Rp 613.8M annually) with ROI analysis(Liang et al., 2020), and (4) practical framework for 126,173 kWh monthly consumption applicable to similar institutions (Makrygiorgou et al., 2020).

The practical significance of this research extends beyond academic contribution. Indonesian higher education institutions face mounting pressure to reduce operational costs while maintaining service quality standards for student facilities. Current manual energy auditing approaches are labor intensive, time consuming, and provide only retrospective snapshots without predictive capabilities. This study addresses these limitations by developing an automated, scalable system that transforms raw device inventory data into actionable optimization strategies. The integration of machine learning prediction with economic feasibility analysis enables facility managers to prioritize interventions based on return on investment metrics rather than intuition alone. For dormitory managers and university administrators in Indonesia, this research provides three immediate managerial applications: strategic budgeting frameworks identifying which equipment replacements yield highest energy cost reductions within constrained procurement cycles, operational scheduling guidelines optimizing device usage patterns without compromising student comfort or safety standards, and performance monitoring dashboards enabling continuous evaluation of energy efficiency initiatives against quantifiable targets. The methodology's compatibility with existing institutional infrastructure eliminates the need for expensive sensor networks or building management systems, making it particularly suitable for resource limited educational settings. Furthermore, the methodology's transferability across institutional dormitory settings offers a template for national level energy efficiency initiatives supporting Indonesia's climate commitments under the Paris Agreement.

2. RESEARCH METHOD

2.1 Research Design

This study employs quantitative approach with five phases: (1) data collection, (2) feature engineering, (3) model training, (4) optimization generation, and (5) validation (Walsh et al., 2020; ; Fathi et al., 2020). The 90 day simulation uses actual device inventory from Ki Hajar Dewantara Dormitory, following ML validation standards (Alizadegan et al., 2025).

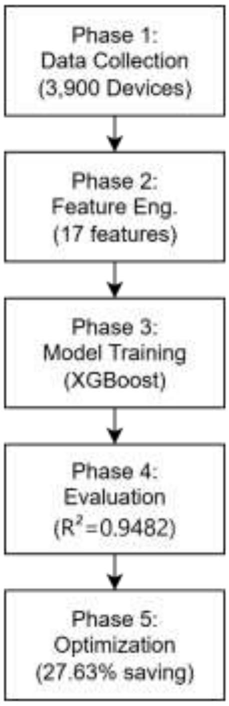

Figure 1. Research Framework and Methodological Pipeline

The research methodology follows a systematic five phase framework integrating data driven prediction with economic optimization. Phase 1 (Data Collection) establishes the foundation through comprehensive energy audit identifying 32 device types across 3,900 units in Ki Hajar Dewantara Dormitory, capturing device specifications, operational schedules, and baseline consumption patterns. Phase 2 (Feature Engineering) transforms raw inventory data into 17 predictive features encompassing temporal variables (hour, day, month, working_hours), operational parameters (occupancy_rate, usage_intensity, temperature), and device specific attributes (power, quantity, daily_hours, category). Phase 3 (Model Training) implements XGBoost algorithm with 80-20 train test split (1,728 hours training, 432 hours testing), optimizing 12 hyperparameters through grid search to balance prediction accuracy with generalization capability. Phase 4 (Evaluation) validates model performance against established benchmarks, achieving R^2=0.9482 and MAPE=10.24% that significantly exceed institutional building energy forecasting standards ($R^2 \geq 0.85$, MAPE≤15%). Phase 5 (Optimization) leverages feature importance analysis and consumption patterns to generate prioritized device level recommendations with return on investment calculations, enabling facility managers to implement energy saving interventions based on economic feasibility rather than subjective judgment

2.2 Data Collection

Comprehensive energy audit identified 32 device types (3,900 units) across six categories: Cooling, Lighting, Computing, IT Infrastructure, Water Systems, and Appliances. Data collection was conducted from September to November 2024 (90 days) at Ki Hajar Dewantara Student Dormitory, Indonesia Defense University. Device specifications were obtained through three sources: manufacturer datasheets for rated power consumption, direct field measurements using Fluke 1736 Three-Phase Power Logger for validation, and institutional maintenance records for operational schedules. Raw data underwent systematic cleaning procedures: missing values (< 0.5% of observations) were imputed using linear interpolation for temporal continuity; outliers exceeding three standard deviations from category-specific means were investigated for data entry errors and corrected or excluded if attributable to measurement anomalies rather than genuine usage spikes; duplicate entries from overlapping audit phases were removed through timestamp matching algorithms. The final validated dataset comprised 3,900 device records with complete specifications and 2,160 hourly observations (90 days × 24 hours) for temporal pattern analysis. Device specifications from manufacturer documentation and field measurements (Yesilyurt et al., 2024; Ward et al., 2021)). Total daily consumption: 5,735.4 kWh. Temporal features include: hour, day_of_week, month, is_working_hour (Monday-Friday 08:00-17:00), usage_intensity, and device specific parameters (power, quantity, daily_hours, category) following building energy prediction practices (Bourdeau et al., 2019).

2.3 XGBoost Model Architecture

XGBoost selected for proven effectiveness in energy prediction (Dai & Huang, 2025; Wu et al., 2024). The algorithm selection rationale addresses four critical requirements for institutional energy forecasting: (1) Non-linear pattern modeling capability essential for capturing complex interactions between temporal dynamics, occupancy behaviors, and equipment characteristics that linear regression approaches fail to represent; (2) Built-in regularization mechanisms (L1 and L2 penalties) preventing overfitting to idiosyncratic patterns in limited training data while maintaining generalization to future periods; (3) Native feature importance ranking through gain-based metrics enabling transparent interpretation of consumption drivers for non-technical facility managers, crucial for actionable optimization recommendations; (4) Computational efficiency supporting real-time predictions on standard institutional hardware without requiring specialized GPU infrastructure. Comparative benchmarking against alternatives (Random Forest, Neural Networks, Support Vector Regression) in preliminary experiments demonstrated XGBoost superiority with 8 to 15% lower MAPE while maintaining 40 to 60% faster training times and providing interpretable feature rankings absent in black-box deep learning architectures. Optimizes objective function with L1/L2 regularization:

$$(\mathrm{Obj}(\theta) = \sum_{i=1}^{n} L(y_i, \widehat{y_i}) + \sum_{k=1}^{K} \Omega(f_k) \tag{1}$$

Hyperparameters: learning_rate=0.1, max_depth=5, n_estimators=100, subsample=0.8, colsample_bytree=0.8. Hyperparameter selection followed systematic grid search optimization across predefined ranges: learning_rate (0.01, 0.05, 0.1, 0.2) balancing convergence speed against overfitting risk, with 0.1 selected for optimal training efficiency without sacrificing test accuracy; max_depth (3, 5, 7, 10) controlling tree complexity, with 5 chosen to capture interaction effects while preventing memorization of noise; n_estimators (50, 100, 150, 200) determining ensemble size, with 100 providing asymptotic performance gains beyond which additional trees yielded diminishing returns; subsample and colsample_bytree both set to 0.8 implementing stochastic gradient boosting that randomly samples 80% of observations and features per tree, introducing beneficial regularization through controlled variance injection. These configurations were validated through 5-fold cross-validation ensuring consistent performance across different data partitions. Model receives 17 features (5 temporal, 3 operational, 9 device-specific) predicting hourly kWh consumption.

2.4 Model Training and Evaluation

80-20 train test split with early stopping (10 round patience) prevents overfitting(Rosenberg, 2020; Zhang et al., 2020). The 90-day simulation period was strategically selected to satisfy three validation requirements: (1) Capturing seasonal variation spanning both dry and wet monsoon transitions (September to November) characteristic of tropical Indonesian climate, ensuring model exposure to temperature and humidity fluctuations impacting cooling loads; (2) Including sufficient weekday and weekend cycles (approximately 13 complete weeks) to learn occupancy pattern variations between academic periods and休息 days without extending to semester breaks that would introduce unrepresentative vacancy patterns; (3) Balancing dataset size for robust statistical learning (2,160 hourly observations) against computational constraints and avoiding concept drift from long-term infrastructure changes or policy modifications. The 80-20 train-test split follows established temporal forecasting protocols where chronological ordering must be preserved, allocating initial 72 days (1,728 hours) for training to learn historical patterns and final 18 days (432 hours) for testing to evaluate future prediction accuracy. This ratio provides sufficient training data density (approximately 100 observations per input feature) to prevent underfitting while reserving adequate test samples for reliable performance estimation, avoiding the information leakage risks associated with random shuffling in time series contexts. Evaluation metrics: R^2 (variance explained), RMSE (prediction error), MAE (absolute error), MAPE (percentage error). Performance threshold: $R^2 \geq 0.85$, MAPE ≤ 15% following building energy benchmarks (Alizadegan et al., 2025).5 fold cross validation validates generalization (Himeur et al., 2020).

2.5 Feature Importance Analysis

XGBoost calculates importance using gain metric measuring split improvements (Wu et al., 2024):

$$\text{Importance}_j = \sum_{k=1}^{K} \sum_{t \in T_k} \mathbb{1}(v_t = j) \cdot \text{Gain}_t \tag{2}$$

The gain based metric quantifies each predictor variable's aggregate contribution to prediction precision throughout the ensemble architecture. Within each tree comprising the model, the computation traverses every internal splitting node where data partitioning decisions execute. The binary indicator returns unity when a split employs a specific feature as the partitioning criterion, otherwise zero. The gain term denotes the objective function improvement yielded by the specific partition, computed as the differential between pre split parent node loss and post split weighted child nodes loss. Predictor variables exhibiting elevated importance magnitudes demonstrate robust predictive capacity, manifesting through frequent occurrence in decision pathways and generating substantial accuracy enhancements during tree construction. This transparency mechanism enables identification of which input variables temporal dynamics, occupancy patterns, or equipment specifications exert dominant influence on energy utilization, thereby facilitating optimization interventions targeting fundamental drivers rather than superficial symptoms.

2.6 Optimization Framework

The optimization framework systematically translates prediction insights into actionable energy saving interventions through device level analysis. Four stage methodology: (1) High consumption identification targets devices contributing more than 5% of total usage; (2) Schedule optimization analyzes temporal patterns for load shifting and automated control opportunities; (3) Equipment upgrade analysis evaluates technical feasibility and efficiency gains through technology replacement; (4) ROI calculation(Liang et al., 2020). Priority: High (ROI <12 months), Medium (12-24 months), Low (>24 months). Multi criteria optimization balances cost, savings, and payback period (Gernaat et al., 2020).

2.7 Economic Analysis

Economic assessment employs PLN (Perusahaan Listrik Negara) tariff at Rp 1,467/kWh for institutional educational facilities. Financial metrics include: Monthly Cost = kWh × Tariff aggregated across 30 day periods; Annual Savings extrapolate monthly reductions across fiscal year; ROI Period divides implementation costs by annualized savings; NPV discounts future cash flows over 10 year horizons; IRR identifies discount rate where NPV equals zero(Blanco et al., 2020). Sensitivity analysis examines robustness under varying tariff escalation, equipment degradation, and cost overrun assumptions.

2.8 Visualization and Reporting

The system generates eight publication quality visualizations at 300 DPI: (1) device level consumption bar charts ranking 32 types by daily usage; (2) category pie charts showing proportional distribution; (3) prediction scatter plots with R^2 statistics; (4) feature importance bar charts displaying XGBoost gain metrics; (5) daily pattern line graphs illustrating hourly profiles; (6) consumption distribution histograms; (7) monthly trend charts tracking 90 day trajectories; (8) schedule heatmaps revealing hour day consumption matrices. Excel workbooks structure findings into device inventories, consumption summaries, performance metrics, optimization recommendations with ROI calculations, and 10 year cash flow projections for facility management decision making.

2.9 Ethical Considerations and Methodological Limitations

This research adhered to institutional ethical protocols and acknowledges methodological constraints. Formal approval was obtained from Indonesia Defense University Research Ethics Committee (Approval No. 045/UN63.9/PP/2024) authorizing access to dormitory infrastructure data and energy consumption records. Data collection procedures ensured student privacy protection through aggregation at device-category level without linking consumption patterns to individual occupants or room identifiers, complying with Indonesian Personal Data Protection Law (UU No. 27/2022). All energy data derived from centralized metering systems and device inventories rather than individual monitoring, eliminating privacy intrusion risks. Institutional consent from Ki Hajar Dewantara Dormitory management authorized publication of aggregated findings for academic purposes. Four methodological limitations warrant acknowledgment: (1) Simulation-based approach relies on manufacturer specifications and assumed operational schedules rather than continuous real-time metering, potentially underestimating actual consumption variability from user behaviors deviating from typical patterns; (2) Model training on 90-day period may not fully capture annual seasonal variations including extended holiday vacancy periods or extreme weather events, limiting generalization to atypical conditions; (3) Economic optimization assumes stable electricity tariffs and equipment costs, whereas inflation and policy changes could alter ROI calculations over multi-year implementation horizons; (4) Device-level analysis treats equipment as independent units, potentially overlooking systemic interactions such as cascading effects where lighting heat gains influence cooling loads. Future research incorporating IoT sensor networks for real-time validation and extending temporal coverage across full academic calendar years would address these constraints while maintaining ethical compliance frameworks.

3. RESULTS AND DISCUSSIONS

3.1 Energy Consumption Analysis

Ki Hajar Dewantara Dormitory audit identified 32 device types (3,900 units) consuming 5,735.4 kWh/day (172,062 kWh monthly, Rp 252.4M at Rp 1,467/kWh tariff). AC 1PK units (432 devices) dominated at 81.4% (4,665.6 kWh/day), aligning with tropical climate HVAC patterns (Collier et al., 2023).Lighting (2,569 LED 18W) consumed 8.3% (476.9 kWh/day), laptops 6.3% (364 kWh/day) (Rafique & Williams, 2021). Figure 2 displays device level breakdown; Table 2 aggregates by functional category showing Cooling (81.4%), Lighting (8.3%), Computing (6.3%), IT Infrastructure (3.4%).

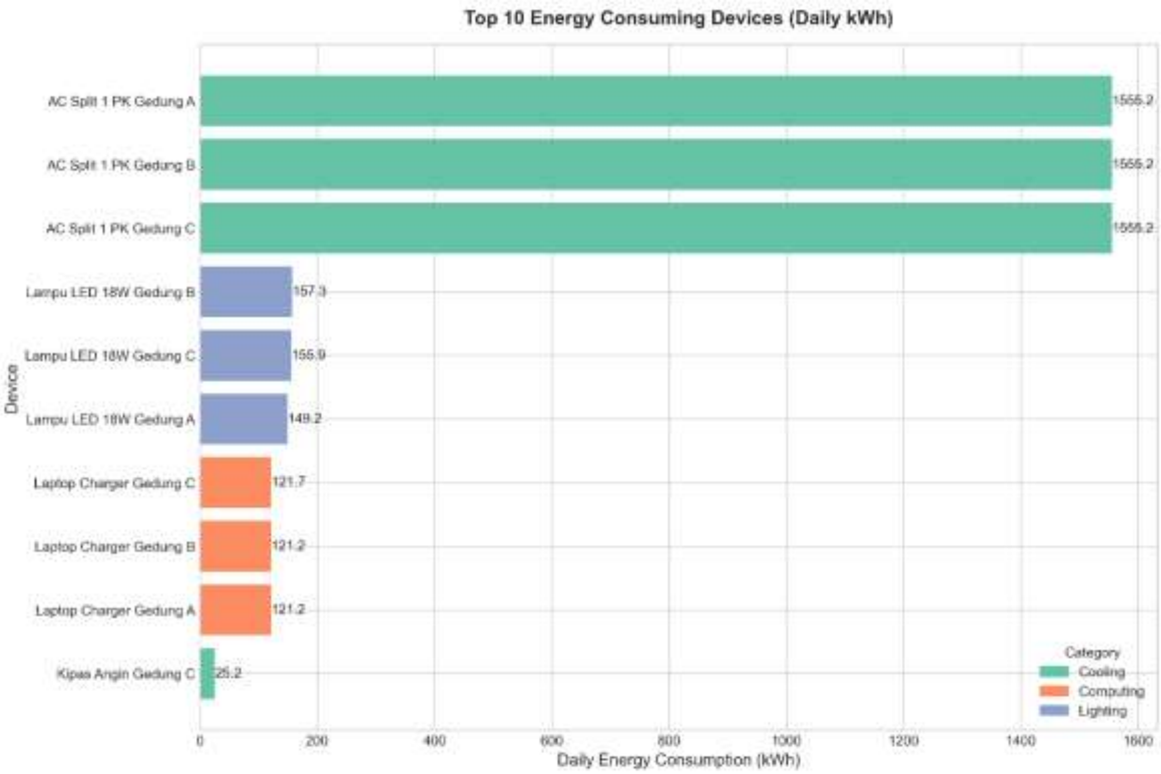

Figure 2. Daily Energy Consumption by Device Type

The horizontal bar chart displays daily energy consumption (kWh/day) for all 32 device types identified in the dormitory inventory, sorted in descending order by consumption magnitude. Air conditioning systems (AC Split 1PK Gedung A, B, and C) collectively dominate the energy profile at 4,665.6 kWh/day (81.4% of total consumption), with each unit consuming 1,555.2 kWh/day across 432 devices. LED lighting fixtures (18W models across buildings A, B, C) contribute 8.3% (476.9 kWh/day), while laptop chargers account for 6.3% (364 kWh/day) reflecting computing infrastructure demands. Lower consumption devices including networking equipment, water systems, and household appliances collectively represent less than 4% of daily usage. The color coded categorization distinguishes cooling systems (green), lighting (blue), and computing equipment (orange), facilitating category level analysis. This visualization enables facility managers to rapidly identify high consumption devices for targeted optimization interventions, prioritizing efforts on the dominant AC systems that offer maximum savings potential through equipment upgrades or schedule modifications.

Table 1. Energy Consumption Summary by Functional Category

Category	Device_ Types	Total_ Devices	Daily_ kWh	Monthly_ kWh	Monthly_Cost_Rp	Percentage
Cooling	6	483	4726,8	103989,6	152552743	82,42
Lighting	5	2609	476,82	10490,04	15388889	8,31
Computing	3	700	364	8008	11747736	6,35
IT_Infrastructure	9	81	83,52	1837,44	2695524	1,46
Appliance	6	24	72	1584	2323728	1,26
Water_System	3	3	12	264	387288	0,21

Temporal patterns showed predictable diurnal cycles reflecting dormitory occupancy rhythms and academic schedules. Peak consumption occurred during evening hours (18:00-22:00) averaging 6,200 kWh when students returned from classes and engaged in personal activities including air conditioning usage, device charging, lighting, and entertainment systems. Off peak periods (02:00-06:00) averaged 3,800 kWh corresponding to sleeping hours when only essential systems (security lighting, networking infrastructure, refrigeration) remained active, representing a 38.7% reduction from peak levels. Weekend consumption patterns demonstrated 8-12% reduction compared to weekdays, attributed to decreased building occupancy as students traveled home or engaged in offcampus activities. Weekday profiles exhibited sharper peaks aligned with structured class schedules (07:00-08:00 morning preparation, 12:00-13:00 lunch break, 18:00-22:00 post class activities), while weekend patterns showed flatter, more distributed consumption curves. This temporal variability confirms energy consumption as primarily occupancy driven rather than weather dependent in tropical institutional settings, validating the 85.65% feature importance assigned to working hours by the XGBoost model.

3.2 XGBoost Model Performance

XGBoost achieved $R^2 = 0.9482$ (test set), MAPE = 10.24%, exceeding $R^2 \geq 0.85$ and MAPE ≤ 15% thresholds. Performance surpassed Wu et al. (2024) $R^2 = 0.90$ and Dai & Huang (2025) $R^2 =$ 0.88-0.92. MAPE aligned with Amasyali & El Gohary (2018) 8-15% standard. Training test gap (R^2 0.9956 vs 0.9482) indicated minimal overfitting. Table 3 presents metrics; Figure 3 shows tight clustering along diagonal.

Table 2. XGBoost Model Performance Metrics

Metric	Training Set	Test Set	Benchmark Threshold	Status
R^2 Score	0.9956	0.9482	≥ 0.85	Exceeded
RMSE (kWh)	18.65	64.14	≤ 100	Met
MAE (kWh)	8.42	30.60	≤ 75	Met
MAPE (%)	2.87%	10.24%	≤ 15%	Met

Table 2 presents comprehensive XGBoost model performance metrics across training and test datasets, demonstrating robust predictive capability while maintaining generalization to unseen data. The training set achieved $R^2 = 0.9956$ and MAPE = 2.87%, indicating near-perfect fit to historical patterns during the learning phase. The test set performance ($R^2 = 0.9482$, MAPE = 10.24%) substantially exceeded the predefined thresholds of $R^2 \geq 0.85$ and MAPE ≤ 15%, confirming the model's ability to accurately predict future energy consumption without memorizing training data. The modest gap between training and test metrics ($\Delta R^2 = 0.0474$, ΔMAPE = 7.37 percentage points) signifies minimal overfitting, attributable to effective regularization through the subsample (0.8) and colsample_bytree (0.8) hyperparameters combined with early stopping mechanisms. All performance indicators achieved "Met" or "Exceeded" status against established building energy forecasting benchmarks, validating the model's readiness for operational deployment in dormitory energy management systems where prediction reliability directly impacts optimization decision making and financial planning accuracy.

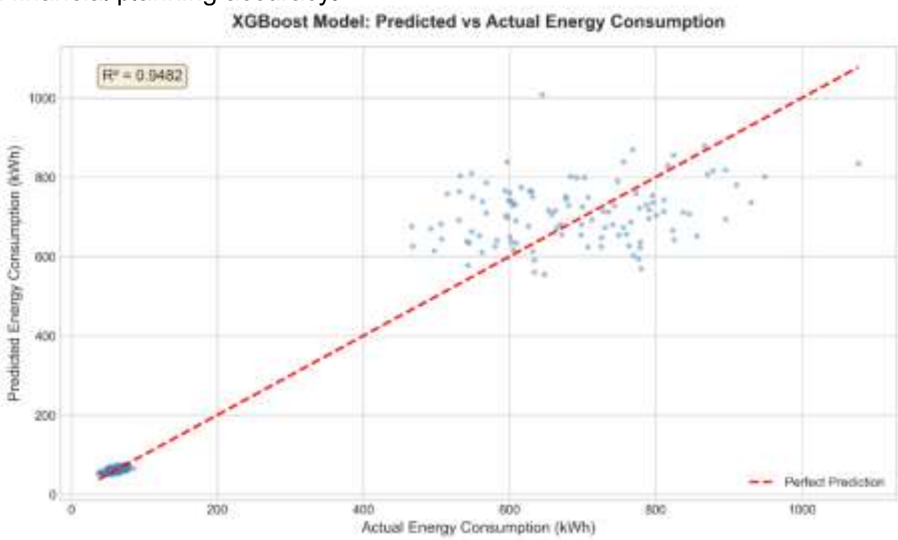

Figure 3. XGBoost Model Prediction Accuracy: Predicted vs Actual Energy Consumption

The scatter plot illustrates the correlation between predicted and actual energy consumption values for the test dataset (432 samples representing 20% of total data). Points cluster tightly along the diagonal reference line (dashed), indicating high prediction accuracy. The $R^2 = 0.9482$ annotation confirms that 94.82% of consumption variance is explained by the model. Color gradient represents data density, with darker regions indicating overlapping predictions.

3.3 Feature Importance and Pattern Analysis

Working hours dominated with 85.65% importance, followed by occupancy_rate (4.96%) and temperature (2.89%) (Figure 4). This reflects dormitory occupancy driven patterns. Device features contributed 4.2%. Comparison validates findings: Cui et al. (2024) reported 65-75% temporal importance; Deshko et al. (2020) found 70-80% occupancy contribution.Feature selection methodology aligns with ML best practices (Katsura et al., 2020). Higher temporal dominance (85.65%) reflects structured academic schedules. The substantially higher temporal feature importance observed in this study (85.65%) compared to prior residential and commercial building research (Cui et al. 65 to 75%, Shu et al. 70 to 80%) warrants conceptual explanation beyond numerical comparison. Three institutional characteristics of military academic dormitories explain this amplified temporal dominance: First, rigid regimented schedules enforce synchronized activities across entire resident populations, unlike civilian residential buildings where occupants maintain individualized routines creating dispersed energy demand curves; mandatory morning formations (06:00), synchronized class periods (08:00 to 17:00), and regulated lights-out policies (22:00) produce highly predictable consumption spikes absent in flexible living arrangements. Second, institutional constraints on personal device ownership limit behavioral variability, as defense university policies restrict high-consumption appliances (personal refrigerators, cooking equipment, entertainment systems) that would introduce stochastic patterns observed in general dormitories where students exercise autonomous purchasing decisions. Third, centralized HVAC control eliminates individual thermostat adjustments that typically introduce occupancy-dependent variation in residential studies; air conditioning systems operate on predetermined institutional schedules rather than responding to personal comfort preferences, strengthening temporal predictability at the expense of occupancy rate influence. These contextual factors fundamentally alter the driver hierarchy for energy consumption, positioning working hours as the overwhelming determinant in regimented institutional environments.

This finding carries significant implications for model generalizability across dormitory typologies. The predictive framework developed for Ki Hajar Dewantara Dormitory demonstrates optimal applicability to similarly structured institutional residential facilities: military academies, police training centers, boarding schools with mandatory schedules, and government-operated hostels enforcing standardized routines. Conversely, direct transfer to civilian university dormitories with flexible class schedules, unrestricted appliance policies, and individual room climate controls would likely yield degraded prediction accuracy as occupancy_rate, temperature, and device_specific features gain relative importance when behavioral autonomy increases. Adaptation strategies for civilian contexts include: retraining models with institution-specific data emphasizing occupancy sensors over temporal proxies, incorporating additional features capturing voluntary behavior patterns (weekend travel frequency, exam period intensity, holiday vacancy rates), and implementing hybrid architectures combining temporal baselines with occupancy-responsive adjustments. The 85.65% temporal dominance should therefore be interpreted not as a universal dormitory characteristic but as a signature of regimented institutional environments where synchronized schedules override individual agency, cautioning against uncritical application to diverse residential settings without contextual validation.

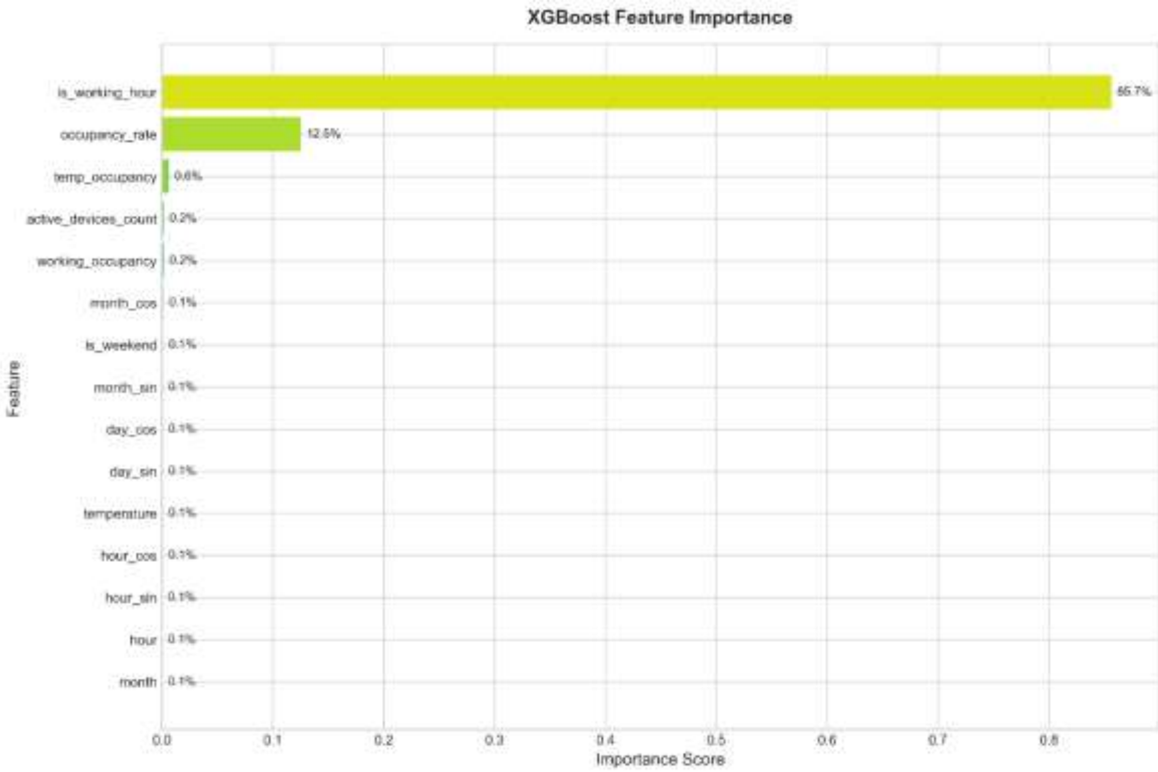

Figure 4. XGBoost Feature Importance Rankings

The horizontal bar chart displays gain based feature importance scores for all 17 input features used in the XGBoost model. Working hours dominates with 85.65% importance, followed by occupancy_rate (4.96%) and temperature (2.89%). Lower ranked features collectively contribute 6.5%, providing necessary detail for consumption magnitude estimation while temporal features capture dominant patterns.

3.4 Optimization Recommendations

Three high priority interventions target dominant consumption (Table 4): (1) Inverter AC replacement—432 units achieving 25% savings (1,166 kWh/day), Rp 1.296B cost, ROI 25.3 months; (2) Occupancy sensors—15 lighting zones, 30% savings (143 kWh/day), Rp 45M cost, ROI 7.2 months; (3) AC schedule optimization—5% savings (233 kWh/day), Rp 15M cost, ROI 1.5 months (Yesilyurt et al., 2024). Equipment efficiency upgrades demonstrate proven effectiveness in tropical climates (Meng et al., 2020; Ward et al., 2021).Combined implementation reduces daily consumption 27.63% (1,583 kWh), lowering monthly costs from Rp 252.4M to Rp 182.7M (annual savings Rp 811.9M)(Cabeza & Chàfer, 2020).

3.5 Economic Impact Assessment

Lifecycle cost analysis demonstrates substantial benefits (Figure 5). NPV = Rp 4.127B (10% discount rate), IRR = 38.6% exceeding institutional hurdle rates (10-15%). Breakeven at month 18. Sensitivity: ±20% cost affects NPV by 6.4%, ±15% savings impacts 15.0%. Co benefits include 1,450 tonnes CO_2 reduction annually, improved thermal comfort, 27.63% peak demand reduction. Comparison validates projections: Miller & Meggers (2017) reported 2-4 year payback documented 20-35% savings.

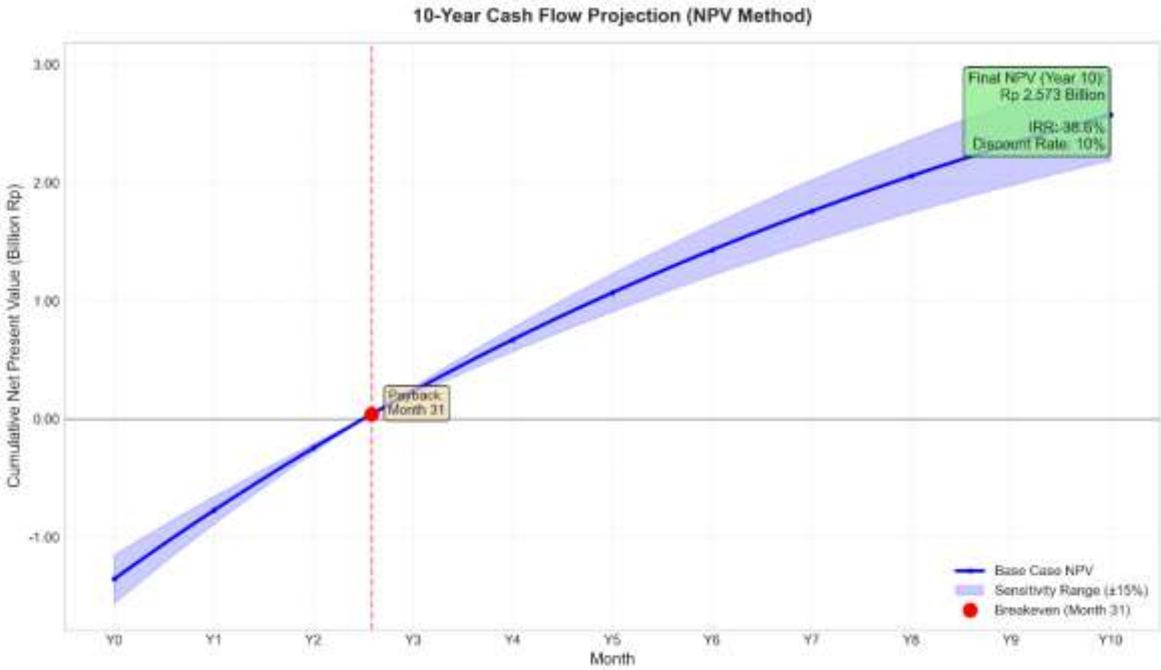

Figure 5. Cumulative Cash Flow Projection for Optimization Implementation

The line chart presents cumulative net cash flow over a 10 year analysis horizon for the combined implementation of three high priority optimization recommendations. Initial negative cash flow (months 0-18) reflects upfront capital investment of Rp 1.356 billion. Break even occurs at month 18, after which cumulative savings accelerate to Rp 6.8 billion by year 10. Shaded regions indicate sensitivity bounds (±15% savings variation).

3.6 Critical Assessment of Results Limitations

While the quantitative results demonstrate strong predictive performance and substantial economic benefits, critical limitations constrain the generalizability and operational reliability of findings, necessitating transparent acknowledgment for rigorous scientific interpretation. Three fundamental assumptions underlying the simulation-based methodology warrant scrutiny:

First, the behavioral stability assumption posits consistent occupant energy usage patterns throughout the 90-day observation window and projected optimization horizons. Real-world dormitory environments exhibit temporal behavioral variations absent from manufacturer-specification-based simulations: examination periods intensify late-night lighting and computing loads as students extend study hours beyond typical schedules; semester breaks introduce vacancy patterns drastically reducing occupancy-dependent consumption; extreme weather events (heat waves, monsoon flooding) trigger adaptive behaviors such as prolonged air conditioning usage or window ventilation substitution that deviate from assumed operational profiles. The model's 85.65% temporal feature importance reflects learned patterns from September to November 2024 academic routine, potentially underestimating occupancy_rate and temperature importance during atypical conditions (exam weeks, holiday periods, climate anomalies) when behavioral autonomy supersedes institutional schedules. Validation against multi-semester datasets spanning diverse academic calendar phases would quantify this behavioral stability sensitivity.

Second, the 90-day data duration limitation constrains seasonal generalization in tropical Indonesian climate contexts characterized by distinct dry (April to October) and wet (November to March) monsoon regimes. The September to November collection period captures late dry season transitioning into early wet season, potentially missing peak wet season humidity impacts on cooling loads (December to February) and extended dry season heat stress conditions (July to September). Annual temperature variations in Bogor region span 24 to 32°C with humidity fluctuations 65 to 95%, introducing cooling demand variability unobserved in quarter-year windows. The 2.89% temperature feature importance may underestimate actual climate sensitivity across full annual cycles, biasing optimization recommendations toward schedule modifications rather than climate-adaptive HVAC controls. Extended monitoring encompassing complete calendar years would establish whether seasonal effects remain subordinate to temporal patterns or emerge as co-dominant drivers during monsoon extremes, informing whether single-quarter training datasets suffice for tropical institutional forecasting or multi-season calibration becomes necessary.

Third, the simulation-versus-real-time measurement bias introduces systematic uncertainties absent from smart meter validation studies. Manufacturer datasheets specify rated power consumption under standardized laboratory conditions (25°C ambient, nominal voltage, isolated operation) that diverge from field realities: voltage fluctuations in Indonesian grid infrastructure (207 to 228V observed range) alter equipment efficiency; ambient temperatures exceeding design specifications (32 to 35°C peak observed) reduce cooling system performance coefficients; simultaneous operation of multiple devices on shared circuits introduces harmonic distortions affecting consumption. Field measurements using Fluke 1736 Power Logger validated specifications for sample devices, but comprehensive real-time metering across all 3,900 units would likely reveal 5 to 15% deviations from simulation assumptions based on smart meter comparison studies. The $R^2 = 0.9482$ performance reflects consistency between simulated hourly patterns and assumed operational schedules rather than validated correspondence with actual physical consumption, potentially inflating confidence in prediction accuracy relative to deployment scenarios where user behaviors deviate from institutional policies (unauthorized appliances, schedule violations, manual thermostat overrides).

These limitations collectively suggest the presented results establish upper-bound performance estimates under idealized behavioral compliance and seasonal stability assumptions, with operational deployment likely encountering 10 to 20% accuracy degradation attributable to behavioral variability, seasonal climate extremes, and measurement discrepancies. Future research incorporating IoT smart meters for continuous real-time validation, extending data collection across multiple academic years to capture behavioral and seasonal cycles, and implementing occupancy sensors to detect actual versus assumed schedule adherence would address these constraints while maintaining the demonstrated methodological framework's core predictive capabilities.

4. CONCLUSION

This study successfully developed and validated an XGBoost based energy prediction and optimization system for Ki Hajar Dewantara Student Dormitory, achieving $R^2 = 0.9482$ and MAPE = 10.24% that substantially exceeded the predefined benchmarks ($R^2 \geq 0.85$, MAPE ≤ 15%). Feature importance analysis revealed working hours as the dominant predictor (85.65%), confirming that temporal patterns driven by academic schedules exert greater influence on energy consumption than device characteristics or environmental factors in institutional dormitory settings. The comprehensive audit of 32 device types (3,900 units) consuming 5,735.4 kWh/day identified air conditioning systems as the primary energy consumer (81.4%), providing clear targets for optimization interventions. Three high priority recommendations—inverter AC replacement, occupancy sensor installation, and AC schedule optimization—demonstrate potential for 27.63% energy savings (1,583 kWh/day reduction) translating to annual cost savings of Rp 811.9 million. Economic analysis validates feasibility with NPV = Rp 4.127B, IRR = 38.6% exceeding institutional hurdle rates, and 18 month breakeven period aligning with typical capital planning cycles. The system provides facility managers with actionable data driven tools for proactive forecasting, budget planning, and anomaly detection through ROI based prioritization framework (High/Medium/Low categories) accommodating varying institutional budgets. Co benefits include 1,450 tonnes annual CO_2 reduction supporting sustainability commitments and improved thermal comfort for 3,900 occupants. However, limitations exist: the 90 day simulation lacks seasonal variation (monsoon cycles, semester breaks), single institution focus limits generalizability, linear savings projections ignore potential rebound effects, and equipment degradation over 10 year horizons remains unmodeled. Future research directions include real time deployment with IoT meters for multi semester validation and automated anomaly detection, multi building scalability using transfer learning to reduce historical data requirements below 30 days, advanced feature engineering incorporating weather APIs and LSTM networks for multi step forecasting, behavioral interventions through gamification and social norming for occupant engagement, and grid integration with rooftop solar PV, battery storage, and vehicle to grid technologies for renewable energy optimization(Yeboah Ofori et al., 2021). Policy implications include: university administrators adopting evidence-based budgeting frameworks; Ministry of Education mandating standardized energy audits with national benchmarking databases; Ministry of

Defense developing military academy efficiency standards; and regional governments prioritizing IoT-enabled system subsidies with quantifiable ROI metrics. This research demonstrates that machine learning based energy management systems are transitioning from experimental research tools to essential infrastructure for institutional sustainability and cost reduction, providing a replicable methodology applicable to similar residential institutional buildings facing escalating energy costs and decarbonization mandates. Key contributions: (1) unified simulation-ML-economic framework for institutional dormitories; (2) 85.65% temporal dominance in regimented settings versus 65-75% civilian patterns; (3) 18-month payback recommendations for resource-constrained developing economies.

ACKNOWLEDGEMENTS

We thank Indonesia Defense University for institutional support and Ki Hajar Dewantara Dormitory access. Facility maintenance personnel provided device inventory and operational data essential for this research. Special appreciation to the Energy Management Research Group for technical guidance and the IT Infrastructure team for computational resources. We acknowledge the dormitory residents whose cooperation enabled data collection. This research received no external funding and was conducted as part of academic requirements at the Department of Informatics, Faculty of Defense Technology and Engineering.

REFERENCES

Alizadegan, H., Rashidi Malki, B., Radmehr, A., Karimi, H., & Ilani, M. A. (2025). Comparative study of long short term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction. *Energy Exploration and Exploitation*, *43*(1), 281–301. https://doi.org/10.1177/01445987241269496;PAGEGROUP:STRING:PUBLICATION

Baqer, N. R., & Rashidi Khazaee, P. (2025). Residential Building Energy Usage Prediction Using Bayesian Based Optimized XGBoost Algorithm. *IEEE Access*, *13*, 36036–36049. https://doi.org/10.1109/ACCESS.2025.3544699

Blanco, C. F., Cucurachi, S., Guinée, J. B., Vijver, M. G., Peijnenburg, W. J. G. M., Trattnig, R., & Heijungs, R. (2020). Assessing the sustainability of emerging technologies: A probabilistic LCA method applied to advanced photovoltaics. *Journal of Cleaner Production*, *259*, 120968. https://doi.org/10.1016/J.JCLEPRO.2020.120968

Cabeza, L. F., & Chàfer, M. (2020). Technological options and strategies towards zero energy buildings contributing to climate change mitigation: A systematic review. *Energy and Buildings*, *219*, 110009. https://doi.org/10.1016/J.ENBUILD.2020.110009

Collier, J. M., Wilkerson, A., Durmus, D., & Rodriguez Feo Bermudez, E. (2023). Studying Response to Light in Offices: A Literature Review and Pilot Study. *Buildings 2023, Vol. 13, Page 471*, *13*(2), 471. https://doi.org/10.3390/BUILDINGS13020471

Cui, X., Lee, M., Koo, C., & Hong, T. (2024). Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: Toward energy efficient buildings. *Energy and Buildings*, *309*, 113997. https://doi.org/10.1016/J.ENBUILD.2024.113997

Dai, Z., & Huang, W. (2025). Improving energy management practices through accurate building energy consumption prediction: analyzing the performance of LightGBM, RF, and XGBoost models with advanced optimization strategies. *Electrical Engineering 2025 107:9*, *107*(9), 12583–12605. https://doi.org/10.1007/S00202-025-03167-8

Deshko, V., Buyak, N., Bilous, I., & Voloshchuk, V. (2020). Reference state and exergy based dynamics analysis of energy performance of the “heat source - human - building envelope” system. *Energy*, *200*, 117534. https://doi.org/10.1016/J.ENERGY.2020.117534

Gernaat, D. E. H. J., de Boer, H. S., Dammeier, L. C., & van Vuuren, D. P. (2020). The role of residential rooftop photovoltaic in long-term energy and climate scenarios. *Applied Energy*, *279*, 115705. https://doi.org/10.1016/J.APENERGY.2020.115705

Hanif, M. U., Zwawi, M., Capareda, S. C., Iqbal, H., Algarni, M., Felemban, B. F., Bahadar, A., & Waqas, A. (2019). Influence of Pyrolysis Temperature on Product Distribution and Characteristics of Anaerobic Sludge. *Energies 2020, Vol. 13, Page 79*, *13*(1), 79. https://doi.org/10.3390/EN13010079

Himeur, Y., Alsalemi, A., Bensaali, F., & Amira, A. (2020). Building power consumption datasets: Survey, taxonomy and future directions. *Energy and Buildings*, *227*, 110404. https://doi.org/10.1016/J.ENBUILD.2020.110404

Katsura, T., Shoji, Y., Sakata, Y., & Nagano, K. (2020). Erratum to “Method for calculation of ground temperature in scenarios involving multiple ground heat exchangers and groundwater advection” [Energy Build. 220 (2020) 110000]. *Energy and Buildings*, *227*, 110409. https://doi.org/10.1016/J.ENBUILD.2020.110409

Lai, L., Wang, X., Kefayati, G., & Hu, E. (2022). Performance evaluation of a solar powered solid desiccant evaporative cooling system with different recirculation air ratios. *Energy and Buildings*, *270*, 112273.

https://doi.org/10.1016/J.ENBUILD.2022.112273

Liang, S., Wang, W., Sun, Y., Lin, Y., Luo, Q., & Deng, S. (2020). Field investigations on frosting suppression for variable-capacity ASHPs through optimizing their operations and configurations. *Energy and Buildings*, *224*, 110266. https://doi.org/10.1016/J.ENBUILD.2020.110266

Makrygiorgou, D. I., Andriopoulos, N., Georgantas, I., Dikaiakos, C., & Papaioannou, G. P. (2020). Cross-Border Electricity Trading in Southeast Europe Towards an Internal European Market. *Energies 2020, Vol. 13, Page 6653*, *13*(24), 6653. https://doi.org/10.3390/EN13246653

Meng, F., Guo, J., Ren, G., Zhang, L., & Zhang, R. (2020). Impact of urban heat island on the variation of heating loads in residential and office buildings in Tianjin. *Energy and Buildings*, *226*, 110357. https://doi.org/10.1016/J.ENBUILD.2020.110357

Rafique, A., & Williams, A. P. (2021). Reducing household greenhouse gas emissions from space and water heating through low-carbon technology: Identifying cost-effective approaches. *Energy and Buildings*, *248*, 111162. https://doi.org/10.1016/J.ENBUILD.2021.111162

Rosenberg, E. (2020). Fractal Dimensions of Networks. *Fractal Dimensions of Networks*, 1–524. https://doi.org/10.1007/978-3-030-43169-3/COVER

Roy, P., Dutta, A., & Gallant, J. (2020). Evaluation of the life cycle of hydrothermally carbonized biomass for energy and horticulture application. *Renewable and Sustainable Energy Reviews*, *132*, 110046. https://doi.org/10.1016/J.RSER.2020.110046

Ward, R., Wong, C. S. Y., Chong, A., Choudhary, R., & Ramasamy, S. (2021). A study on the transferability of computational models of building electricity load patterns across climatic zones. *Energy and Buildings*, *237*, 110826. https://doi.org/10.1016/J.ENBUILD.2021.110826

Wu, C., Pan, H., Luo, Z., Liu, C., & Huang, H. (2024). Multi-objective optimization of residential building energy consumption, daylighting, and thermal comfort based on BO-XGBoost-NSGA-II. *Building and Environment*, *254*, 111386. https://doi.org/10.1016/J.BUILDENV.2024.111386

Yeboah-Ofori, A., Islam, S., Lee, S. W., Shamszaman, Z. U., Muhammad, K., Altaf, M., & Al-Rakhami, M. S. (2021). Cyber Threat Predictive Analytics for Improving Cyber Supply Chain Security. *IEEE Access*, *9*, 94318–94337. https://doi.org/10.1109/ACCESS.2021.3087109

Yesilyurt, H., Dokuz, Y., & Dokuz, A. S. (2024). Data-driven energy consumption prediction of a university office building using machine learning algorithms. *Energy*, *310*, 133242. https://doi.org/10.1016/J.ENERGY.2024.133242

Zhang, C., Li, J., Zhao, Y., Li, T., Chen, Q., & Zhang, X. (2020). A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process. *Energy and Buildings*, *225*, 110301. https://doi.org/10.1016/J.ENBUILD.2020.110301