Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 11, No. 4, November, Pp. 641-650

Enhancing model trust in load forecasting using counterfactual explanations

Muhammad Syarif Hidayatullah*[1], Wahyul Amien Syafei[1], Tarno[1]
Master Program of Information Systems, Postgraduate School, Universitas Diponegoro, Indonesia[1]

Article Info
Keywords:
Counterfactual Explanations, Electricity Load Forecasting, Random Forest Regression, DiCE Framework, Model Interpretability

Article history:
Received: November 04, 2025
Accepted: February 08, 2026
Published: November 01, 2026

Cite:
M. S. Hidayatullah, Wahyul Amien Syafei, and Tarno, "Enhancing Model Trust in Load Forecasting using Counterfactual Explanations", *KINETIK*, vol. 11, no. 4, Nov. 2026.
https://doi.org/10.22219/kinetik.v11i4.2565

*Corresponding author.
Muhammad Syarif Hidayatullah
E-mail address:
hidayatullahms@students.undip.ac.id

Abstract
This study investigates the effectiveness of Counterfactual Explanations (CE) in improving the interpretability of short-term electricity load forecasting models. The forecasting model was developed using Random Forest Regression (RFR) due to its ability to capture nonlinear multivariate patterns in electricity load data and provide global, descriptive interpretability through feature-importance analysis. However, feature importance offers static and potentially biased explanations, making it insufficient for instance-level interpretability. To address this limitation, the Diverse Counterfactual Explanations (DiCE) framework was integrated to generate multiple "what-if" scenarios that illustrate model-consistent input adjustments capable of shifting forecast outcomes. The study uses hourly electricity load data from Panama collected over 2015–2020, with 2020 excluded during preprocessing due to atypical demand patterns during COVID-19. The model was trained using time-series cross-validation and optimized through grid search. Forecasting performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Counterfactual Explanations were assessed using validity, proximity, and compactness metrics, indicating that most generated scenarios are goal-aligned and remain moderately close to the original instances, while compactness results reflect a feasibility trade-off when large prediction shifts are required. Overall, combining RFR with counterfactual analysis enhances model transparency by bridging global and local interpretability perspectives and provides decision-relevant, scenario-based insights to support transparent and data-driven planning in the energy sector.

1. Introduction

Electricity load forecasting is a critical component in the management of modern energy systems, where accuracy plays a vital role in supporting operational planning, optimizing utility performance, and improving cost efficiency [1]. A reliable forecasting model facilitates collaboration between industries and governments to reduce energy waste, maintain grid stability, and enhance the economic value of power system operations [2].

Recent advancements in Artificial Intelligence (AI) have significantly improved forecasting performance; however, these models often face a trade-off between accuracy and interpretability. In this regard, ensemble learning models have emerged as a promising alternative, offering competitive predictive accuracy without entirely compromising model transparency [3]. Among ensemble methods, Random Forest Regression (RFR) is widely used due to its ability to provide global, descriptive interpretability through feature importance measures [4], [5]. RFR is selected for its robustness in modeling complex nonlinear load dynamics and interactions, offering a practical balance between predictive performance and model transparency relative to traditional statistical approaches (e.g., ARIMA) and more opaque deep learning models (e.g., LSTM) commonly used in load forecasting [6], [7], [8]. Comparative regression studies in energy demand forecasting further report tree-based regressors as competitive alternatives in this domain [9].

Nevertheless, RFR's interpretability remains limited, as it is inherently global and model-centric, often introducing feature bias and failing to provide actionable, instance-level explanations [10], [11]. To address this gap, Explainable AI (XAI) through Counterfactual Explanations (CE) represents an intuitive and actionable framework. CE identifies the smallest changes to input features to shift a model's prediction toward a desired outcome, answering practical "what-if" questions [12], [13]**.** This supports decision-relevant interpretability by enabling stakeholders to evaluate potential actions and multiple intervention scenarios beyond traditional analysis [14], [15], [16], [17], [18].

Despite its potential, most existing studies on CE focus on classification models, with relatively few investigations in time-series forecasting, especially for electricity load data with distinct seasonal characteristics [19], [20]. This research gap highlights the opportunity to expand CE applications to the energy domain. Consequently, this study

integrates RFR with the Diverse Counterfactual Explanations (DiCE) framework to generate multiple counterfactual scenarios, ensuring a balance between predictive performance and multi-level interpretability [21], [22], [23], [24], [25].

This RFR–DiCE configuration balances predictive performance with multi-level interpretability, aligning the model with trust-driven decision-support objectives. Despite the potential of CE, a significant research gap exists, as most studies focus on classification, leaving limited empirical evidence on its application in time-series forecasting, particularly for electricity load data with its distinct seasonal characteristics. This study addresses this gap by combining RFR with a model-agnostic CE framework and evaluating both forecasting accuracy and counterfactual quality.

In line with this, the study aims to enhance forecast transparency and decision relevance. Model reliability is assessed using MAE, RMSE, and MAPE, while the generated counterfactuals are evaluated through validity, proximity, and compactness metrics to measure their effectiveness in achieving desired prediction shifts. Furthermore, a scenario-based analysis of feature modification frequency is conducted to examine realistic intervention strategies in load management. These findings serve as an initial step toward expanding the use of CE in the energy sector.

The main contributions of this study are as follows:

- **Development of a robust RFR-based forecasting model** using hourly load data enriched with weather and calendar features, evaluated through standard error metrics (MAE, RMSE, MAPE).
- **Integration of the DiCE framework** with RFR to generate diverse, actionable "what-if" scenarios for load reduction, quantified via validity, proximity, and compactness metrics.
- **Comparative interpretability analysis** between global feature importance and local counterfactuals, demonstrating how both perspectives provide complementary insights into model behavior.
- **Discussion of practical implications** and the trade-offs in counterfactual quality (validity vs. proximity vs. compactness) to provide a roadmap for decision-oriented interpretability in the energy sector.

These contributions position the RFR–DiCE framework as a practical approach to enhancing transparency and trust in short-term electricity load forecasting.

2. Research Method

This research was conducted through five interrelated stages: data preprocessing, model training, model testing, Counterfactual Explanations generation, and evaluation. Each stage produces outputs that serve as inputs for the subsequent step, ensuring a systematic research workflow, as shown in Figure 1.

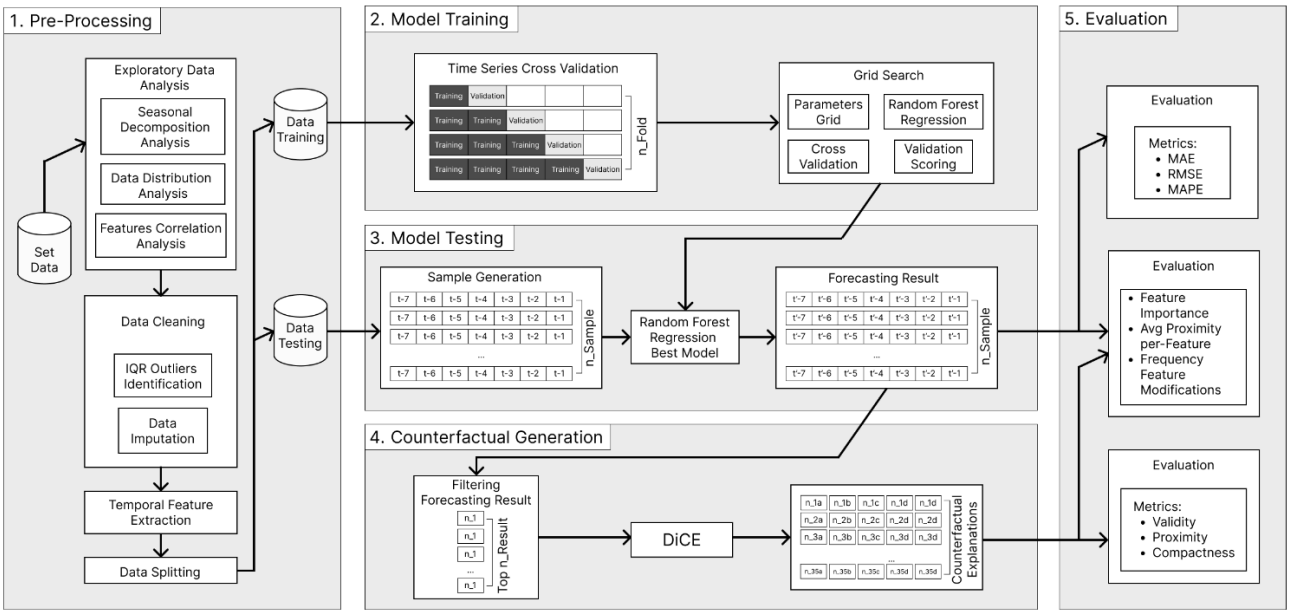


Figure 1. Research Method Consist of Five Interrelated Stages

2.1 Preprocessing Stage

The preprocessing stage ensures data quality and temporal consistency for reliable forecasting and interpretability analysis. The initial dataset comprises hourly electricity load data from Panama (2015–2020), including meteorological and calendar variables [26]. Table 1 provides an illustrative overview of the variables and data structure.

© 2026 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY NC SA license (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Table 1. Variables and Structure of the Hourly Electricity Load Dataset from Panama (2015–2020).

datetime	nat_demand	T2M_toc	QV2M_toc	TQL_toc	W2M_toc	…	Holiday_ID	holiday	school
03/01/15 01.00	970.34	25.86	0.019	0.016	21.851	…	0	0	0
03/01/15 02.00	912.17	25.89	0.019	0.016	22.167	…	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
26/06/20 22.00	1081.56	26.97	0.020	0.057	10.262	…	0	0	1
26/06/20 23.00	1041.62	26.86	0.020	0.064	10.327	…	0	0	1
27/06/20 00.00	1013.56	26.75	0.020	0.064	10.198	…	0	0	1

Following initial inspection, the year 2020 was excluded due to atypical demand patterns associated with the COVID-19 pandemic; thus, data from 2015 to 2019 were utilized. The preprocessing consists of three main components: exploratory analysis, data cleaning, and feature extraction.

2.1.1 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was performed to characterize the dataset structure and variable distributions [27], [28]. First, an additive seasonal decomposition with a 168-hour period (24 hours × 7 days) was applied to separate the load series into trend, seasonal, and residual components. This enabled the capture of daily and weekly patterns while suppressing short-term noise. Second, weekly consumption distributions were visualized using boxplots to compare weekday and weekend patterns and identify potential anomalies. Third, a Pearson correlation analysis was conducted to quantify relationships between variables. These insights guided the selection of relevant input features and helped mitigate multicollinearity, ensuring model efficiency.

2.1.2 Data Cleaning and Feature Extraction

Data cleaning focused on improving the reliability of model inputs. Outliers were identified using the Interquartile Range (IQR) method and treated as missing values. These, along with any existing null values, were handled using hour-of-day mean imputation, which preserves intraday load regularities more effectively than global mean imputation.

Following cleaning, temporal feature extraction was performed to capture recurring periodic patterns. This included extracting calendar features and constructing lag features by shifting historical load values backward. These lag features provide the model with explicit information regarding short-term temporal dependencies. All timestamps were converted to a consistent datetime format and sorted chronologically to maintain time-series integrity.

2.1.3 Data Splitting

The dataset was divided using a chronological (time-ordered) split to prevent information leakage. The training set comprises hourly observations from 2015 to 2017. The independent test set consists of 17,522 observations from 2018 to 2019, providing a two-year evaluation horizon that spans multiple seasonal conditions for a realistic performance assessment.

2.2 Model Training

The Random Forest Regression (RFR) model was trained using Time Series Cross-Validation with an expanding window scheme to preserve temporal order and prevent information leakage. This approach reflects real-world forecasting by incrementally increasing the training set size for each fold. To achieve optimal performance, hyperparameter tuning was performed via Grid Search, evaluating combinations of tree count, maximum depth, and node splitting criteria. Each configuration was assessed using Root Mean Squared Error (RMSE) across the validation folds, ensuring that the final model generalizes effectively to unseen electricity load patterns.

2.3 Model Testing

Model evaluation was conducted on the independent 2018–2019 test set using a structured sampling approach to simulate realistic weekly forecasting. The test data was partitioned into 35 non-overlapping samples, each consisting of seven consecutive days (168 hourly instances), totaling 5,880 unique test instances. Starting points for these windows were randomly selected at 00:00, ensuring no data reuse across samples. This design maintains chronological integrity and prevents data leakage by restricting the model's input to the seven days preceding each forecast, thereby reflecting performance in practical short-term forecasting scenarios.

2.4 Counterfactual Explanations Generation

Local interpretability was enhanced using the Diverse Counterfactual Explanations (DiCE) framework to generate "what-if" scenarios for the RFR model. From the 35 test samples, the instance with the highest forecasted load in each sample was selected as the baseline, simulating strategies for peak load reduction. For each baseline, DiCE generated five counterfactual alternatives with a target of a 20% load reduction, resulting in 175 counterfactual instances for evaluation.

All original input features were provided as candidates for perturbation, with numerical values constrained to the training data's range. In the analysis, a distinction was made between feature types: lagged variables were treated as immutable structural drivers, while exogenous (weather) and calendar features were emphasized as actionable intervention candidates for scenario-based planning, such as demand-response programs.

2.5 Evaluation

Predictive accuracy was assessed using MAE, RMSE, and MAPE. To evaluate Counterfactual Explanations (CE), three metrics—validity, proximity, and inverse compactness—were computed for each of the 175 generated counterfactual instances and subsequently aggregated into sample-level scores for the 35 selected peak instances. This aggregation scheme enables consistent cross-instance comparison while preserving the interpretability of each metric.

Validity indicates whether a counterfactual successfully achieves the predefined target reduction and serves as a success-oriented evaluation criterion commonly adopted in counterfactual analysis [29]. Proximity measures the magnitude (or cost) of feature adjustments required to obtain a valid counterfactual instance. Inverse compactness captures interpretability by quantifying the sparsity of modifications, reflecting how many features must be altered to produce a meaningful output shift, consistent with sparsity-oriented counterfactual evaluation frameworks [30].

Beyond metric-based evaluation, a multi-layered analysis was conducted to further characterize model decision-making and DiCE's manipulation patterns. First, global feature importance identifies the primary predictors (lags, weather, and calendar variables) driving the forecasts. Second, average proximity per feature was reported to assess the stability and realism of modifications; higher mean proximity indicates more conservative adjustments, whereas lower values imply more substantial deviations. Third, feature modification frequency was analyzed to quantify how often each variable acts as a lever for prediction shifts. This combined approach distinguishes between features that are globally influential in the forecasting model and those that are practically relevant for constructing scenario-based counterfactual interpretations.

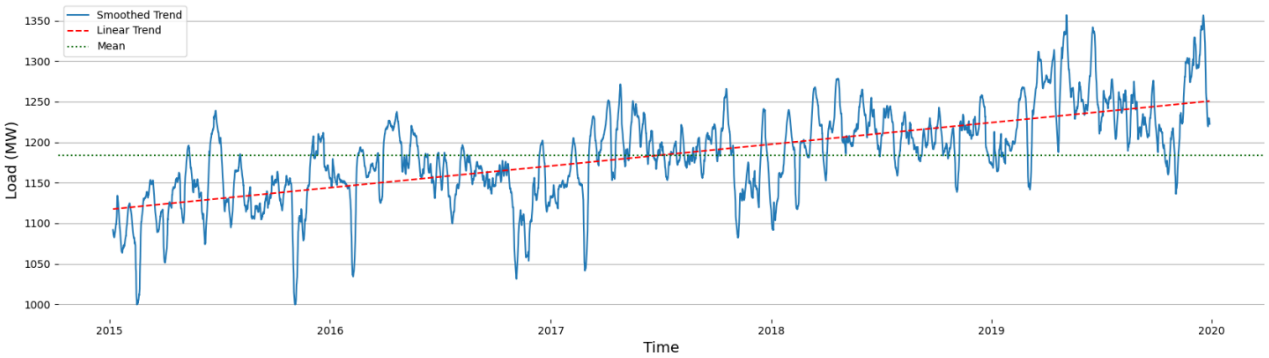
3. Results and Discussion

3.1 Exploratory Data Analysis

Seasonal decomposition of Panama's hourly load (2015–2019) reveals a clear upward trend, increasing from 1,100 MW to approximately 1,300 MW, alongside a stable 168-hour seasonal cycle, as shown in Figure 2. The dominance of the seasonal component suggests that short-term variability is primarily driven by recurring daily and weekly consumption patterns. While residuals are generally centered around zero, occasional extremes indicate idiosyncratic deviations that introduce forecasting uncertainty.

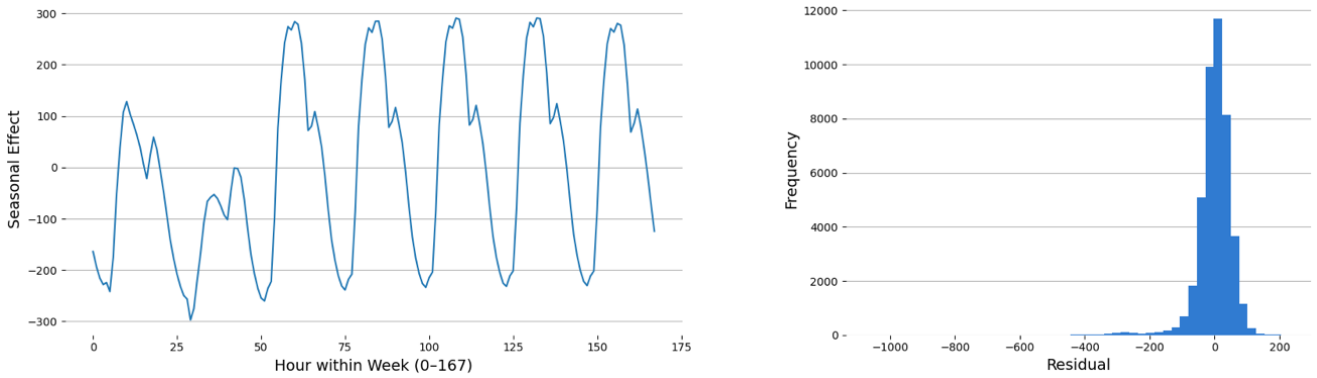
Weekly distribution analysis, as shown in Figure 3, confirms consistent consumption during weekdays (days 0–4), whereas weekends (days 5–6) exhibit lower median loads and more frequent low-end outliers. This nonlinear behavior justifies the use of ensemble-based RFR to capture complex temporal dependencies.

Correlation analysis in Figure 4 identified strong multicollinearity among meteorological variables and redundant holiday indicators. To enhance model efficiency, temperature, humidity, precipitation, and wind features were averaged into representative aggregates (*temp, humid, prec, wind*), and only the *holiday* feature was retained.



(a)

© 2026 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY NC SA license (https://creativecommons.org/licenses/by-nc-sa/4.0/)



(b) (c)

Figure 2. Results of Seasonal Decomposition of Panama's Hourly Electricity Load Data (2015–2019) using an Additive Model with a Seasonal Period of 168 Hours: (a) Smoothed Trend Component Over the Full Period, (b) Average Weekly Seasonal Profile (168-hour cycle) Representing Typical Intraday and Intraweek Variations, and (c) Histogram of Residuals Illustrating the Distribution of Unexplained Variations

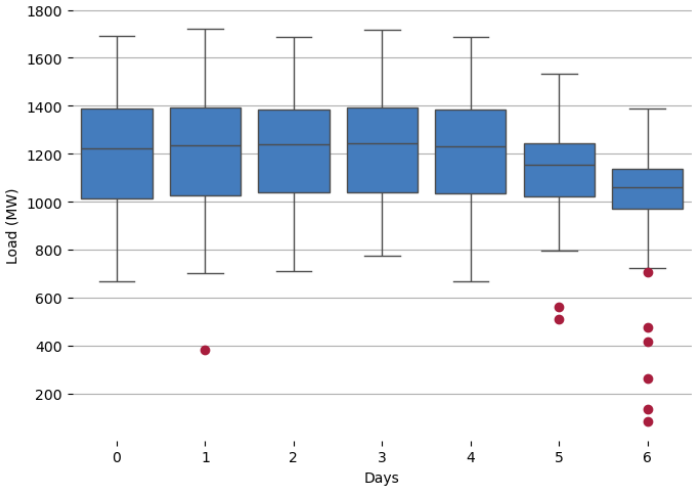


Figure 3. Weekly Distribution (boxplot) of Panama's Hourly Electricity Load Data (2015–2019), with Circular Markers Indicating Outliers

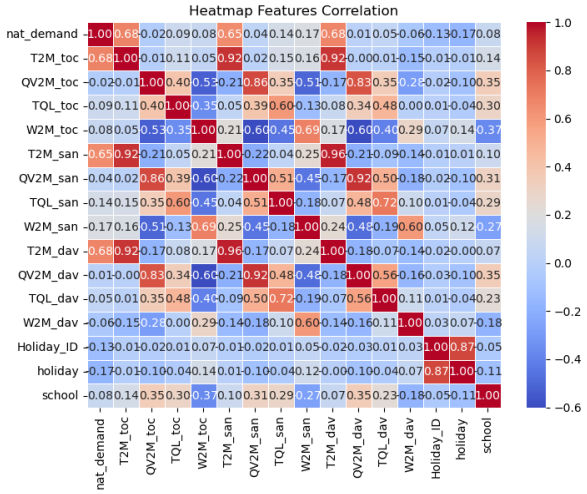


Figure 4. Correlation Heatmap of Input Variables in Panama's Hourly Electricity Load Data (2015–2019), where Values Close to 1 Indicate a Strong Positive Relationship, Values Close to −1 Indicate a Strong Negative Relationship, and Values Near 0 Indicate Weak or Negligible Associations

Based on the Exploratory Data Analysis (EDA) and correlation analysis, the final set of input features used for modelling consists of aggregated meteorological variables—**temperature (temp), humidity (humid), precipitation (prec), and wind speed (wind)**—together with calendar indicators (**holiday** and **school**) and the selected weekly lag features (**lag_168, lag_336, lag_504, and lag_672**). With the dataset consistently structured and the final feature set defined, the next stage focused on developing and evaluating the performance of the Random Forest Regression (RFR) model for short-term electricity load forecasting.

3.2 Electricity Load Forecasting Model Performance

The RFR model demonstrates stable forecasting accuracy across the test period, as summarized in Table 2. With an average MAE of 43.45 MW and MAPE of 3.54%, the model proves reliable for routine operational planning. The slightly higher average RMSE (51.52 MW) indicates sensitivity to occasional peak demand spikes rather than a systematic bias. While this evaluation primarily reflects absolute performance within the defined experimental setting, comparison with recent Random Forest-based short-term load forecasting research suggests that the obtained MAPE is consistent with reported performance levels under comparable methodological configurations [6]. It is important to note that differences in datasets, forecasting horizons, and feature engineering strategies limit strict cross-study comparability. Nevertheless, these findings indicate that the proposed RFR configuration achieves competitive predictive accuracy while enabling additional interpretability analysis through counterfactual explanations.

Table 2. Summary Statistics of Trained Random Forest Regression (RFR) Model Performance on the Evaluated Test Samples of Panama Electricity Forecasting (2018–2019)

	MAE	RMSE	MAPE (%)
Min	25.94	30.96	2.19
Max	75.27	82.48	6.17
Mean	43.45	51.52	3.54
Std	13.41	14.41	1.03

Global interpretation via feature importance, as shown in Figure 5, reveals that historical dependencies are the primary drivers of the forecast. The one-week lag (*lag_168*) emerged as the most dominant predictor (importance ≈0.23), followed by subsequent weekly lags (*lag_672, lag_336, lag_504*). This underscores the strong weekly seasonality inherent in the dataset. In contrast, external factors like *holiday* and *temp* (importance ≈0.06–0.07) provide supplementary insights, while humidity, precipitation, and wind speed show minimal direct impact. These results confirm that the RFR model effectively prioritizes recurring historical cycles, supplemented by critical exogenous effects, to maintain predictive reliability.

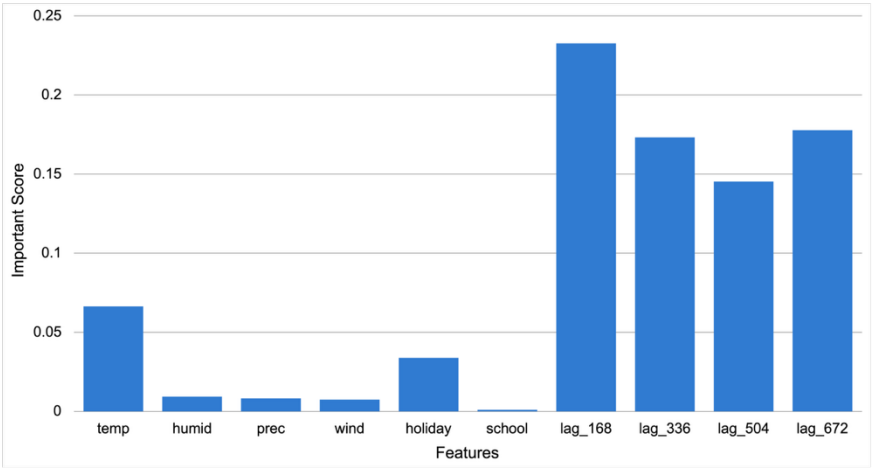


Figure 5. Global Feature Importance Profile of the Trained Random Forest Regression (RFR) Model, Illustrating the Relative Contribution of Each Input feature to Electricity Load Forecasting

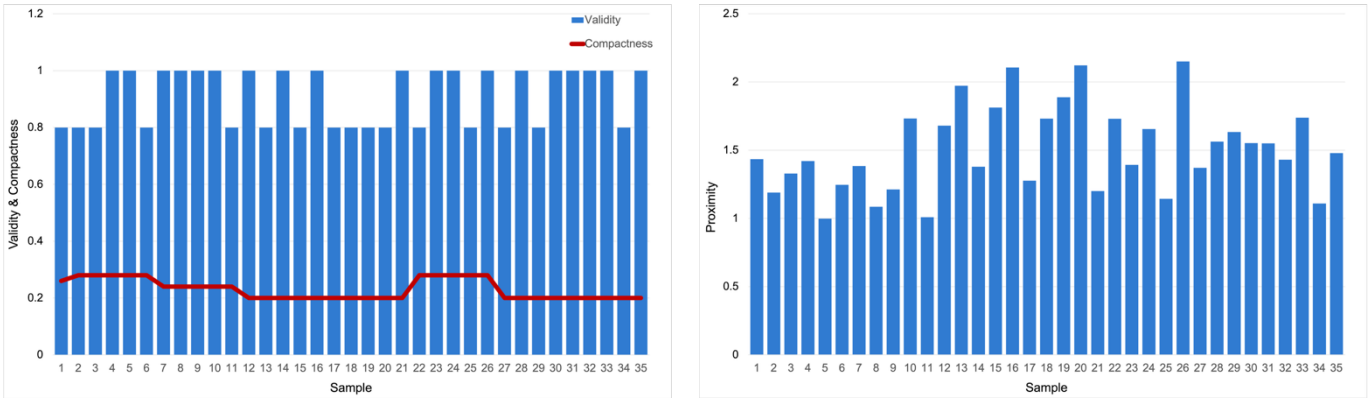
3.3 Counterfactual Explanations Analysis

3.3.1 Quantitative Quality of Counterfactual Explanations

The quality of the Counterfactual Explanations (CE) was first evaluated at the sample level to examine performance consistency across different forecasting contexts, as shown in Figure 6. The bar plots in Figure 6(a) show that validity scores for nearly all 35 samples lie in the high range (0.8–1.0), indicating that DiCE consistently identifies

© 2026 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY NC SA license (https://creativecommons.org/licenses/by-nc-sa/4.0/)

feature modifications that successfully shift predictions toward the desired 20% reduction target. While proximity values in Figure 6(b) exhibit moderate variation (1.0–2.2), they remain within a plausible range without extreme spikes, suggesting that the generated "what-if" scenarios are operationally grounded despite the heterogeneity of electricity load patterns. Conversely, the inverse compactness shown as a line plot in Figure 6(a) exhibits a stable but lower trend (0.2–0.3), confirming that achieving significant load shifts consistently requires coordinated adjustments across multiple features rather than a single variable change.



(a) (b)

Figure 6. Evaluation Results of Counterfactual Explanations Based on Averaged Validity, Proximity, and Compactness Scores Across 35 Test Samples: (a) Dual-axis Plot Showing the Validity and Compactness for Each Sample and (b) Proximity Values Representing the Magnitude of Feature Changes Required to Obtain Valid Counterfactual Instances

To complement this qualitative observation of inter-sample variation, descriptive statistical summaries of validity, proximity, and compactness across all 35 test samples are presented in Table 3 as a basis for quantitative evaluation.

Table 3. Summary Statistics of Counterfactual Quality Metrics Across Evaluated Samples

	Validity	Proximity	Compactness
Min	0.80	0.998	0.20
Max	1.00	2.151	0.30
Average	0.91	1.506	0.23
Median	1.00	1.434	0.20
Std	0.10	0.315	0.04

The average validity of 0.91 (median 1.00) reinforces the framework's reliability in generating functionally valid explanations across the test horizon. The average proximity of 1.506 (std 0.315) reflects the "cost" of these modifications, where the relatively low standard deviation suggests that the magnitude of required changes is fairly consistent across samples. Finally, the average compactness (0.23) highlights a structural characteristic of the model: because a 20% load reduction is a substantial target, the algorithm must explore broader feature combinations to satisfy the constraint. Consequently, these CEs function more effectively as analytical tools for sensitivity analysis rather than as direct operational prescriptions.

3.3.2 Feature Modification Patterns

The stability and realism of the generated counterfactuals are further analyzed through per-feature proximity scores, as presented in Figure 7. Meteorological variables, particularly humidity and temperature, exhibit the highest mean proximity (approximately 0.48–0.50), indicating that DiCE achieves target reductions through minimal and stable adjustments to these inputs. Historical lag features also maintain moderate proximity (0.39–0.45), suggesting they provide a consistent structural basis for the scenarios. In contrast, calendar-related variables like school and holiday show markedly lower proximity, reflecting larger, less stable deviations. This pattern suggests that while weather and lags act as "anchor" features that can be adjusted subtly, calendar features are treated as more rigid or less influential constraints within the local search space.

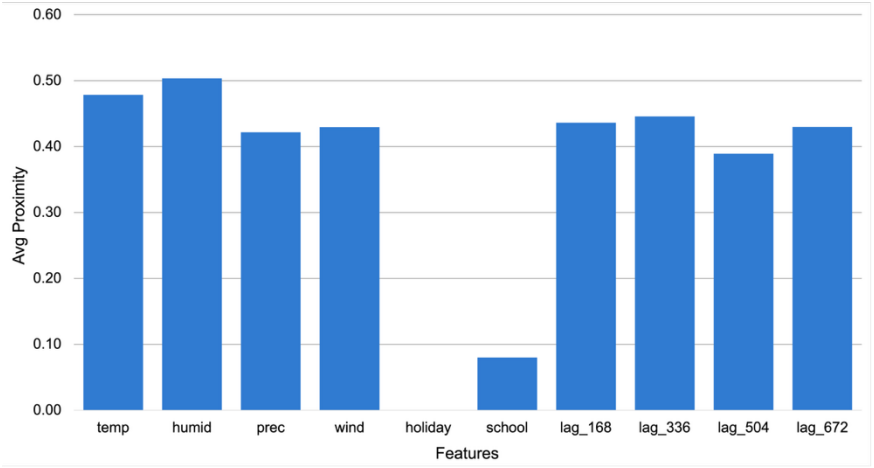


Figure 7. Average Per-feature Proximity Across Generated Counterfactual Instances for Panama Electricity Load Forecasting (2018–2019)

This stability is complemented by the feature modification frequency analysis, presented in Figure 8, which identifies the specific "levers" utilized to reach the 20% reduction target. Meteorological variables (*temp, humid, prec, wind*) are the most frequently altered, appearing in nearly all 175 counterfactual instances. This high frequency, combined with their high proximity, confirms that weather-related adjustments are the most efficient and plausible mechanism for load reduction in the model's logic. Lag features follow with moderate-to-high frequency, while *holiday* is almost never modified. The rare utilization of calendar features, despite their global importance, highlights a critical distinction: these variables serve as contextual markers for the forecast but lack the flexibility required to act as actionable levers for localized "what-if" interventions.

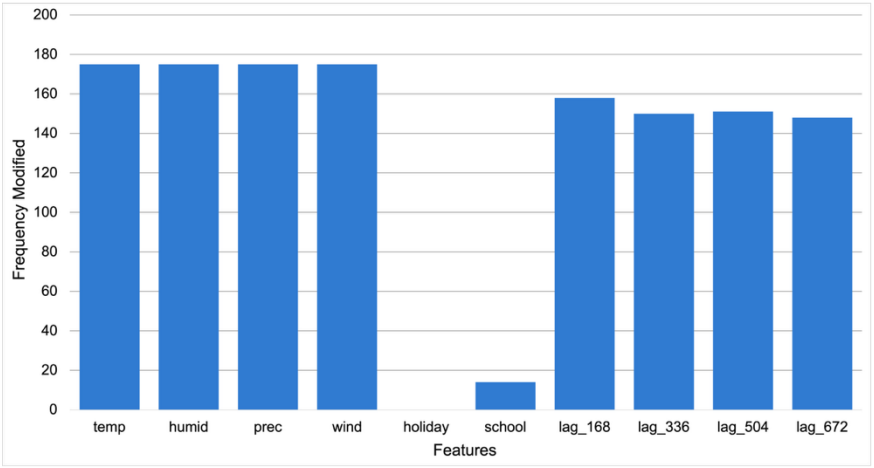


Figure 8. Frequency of Feature Modifications Across All Generated Counterfactual Instances for Panama Electricity Load Forecasting (2018–2019)

Taken together, these analyses demonstrate that the RFR–DiCE framework aligns with the physical reality of electricity demand, where load variability is highly sensitive to environmental conditions and historical consumption cycles. By prioritizing weather and lag features as the primary drivers for counterfactual generation, the model produces scenarios that are not only statistically valid but also operationally interpretable. These findings provide a transparent basis for energy decision-makers to explore how specific, realistic shifts in external conditions—such as demand-response measures tied to temperature—can effectively influence overall load profiles.

3.3.3 Complementary Roles of Feature Importance and Counterfactual Explanations

The interpretability analysis demonstrates that Feature Importance (FI) and Counterfactual Explanations (CE) provide distinct yet complementary perspectives that are essential for a holistic understanding of the forecasting model. Global feature importance establishes the structural foundation of the model by identifying historical lag variables, particularly *lag_168*, as the primary predictive anchors. These features represent the long-term seasonal consistency of electricity consumption. However, the CE analysis reveals a critical "importance-actionability gap" where

© 2026 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY NC SA license (https://creativecommons.org/licenses/by-nc-sa/4.0/)

meteorological variables—despite having lower global importance scores—emerge as the most active instruments for local prediction shifts. While lag features define the baseline forecast, weather variables such as temperature and humidity function as the primary levers for constructing realistic "what-if" scenarios, as they require the smallest magnitude of change to produce significant shifts in the output.

Furthermore, the integration of these two methods clarifies the role of categorical variables like holidays, which possess noticeable global importance but exhibit near-zero flexibility in local counterfactual generation. This distinction is vital for operational decision-making; it confirms that certain influential features act as static contextual markers rather than adjustable intervention points. By synthesizing the global dominance of historical lags with the local agility of environmental variables, this study provides a more transparent and actionable framework. This dual-layered interpretability allows stakeholders to not only identify what drives the load forecast but also to evaluate the feasibility of specific adjustments, thereby bridging the gap between automated forecasting and practical energy management strategies.

4. Conclusion

This study integrated Random Forest Regression (RFR) with the Diverse Counterfactual Explanations (DiCE) framework to enhance decision-level interpretability in short-term electricity load forecasting. The RFR model demonstrated stable predictive performance with an average MAPE of 3.54%. The counterfactual evaluation revealed that while DiCE consistently generates valid "what-if" scenarios (average validity of 0.91), a significant trade-off exists between proximity and compactness. Achieving substantial load reduction often requires coordinated modifications across multiple features, which may constrain immediate operational actionability but provides deep insights into the model's sensitivity.

The primary contribution of this work lies in the synergy between global and local interpretability. Feature Importance identifies the structural "anchors" of the model (historical lags), while Counterfactual Explanations uncover the actionable "levers" (meteorological variables). This complementarity enables decision-relevant, scenario-based interpretability, allowing stakeholders to transcend static feature-based insights and explore plausible alternative scenarios in support of more risk-aware decision-making in power system management.

Despite these contributions, the study is limited by the absence of comparative forecasting baselines and the moderate compactness of some counterfactuals. Future research should prioritize benchmarking against alternative models and refining counterfactual generation with feasibility-oriented constraints. Additionally, incorporating user-centered evaluations with system operators could provide qualitative insights into the perceived utility and clarity of these scenarios in real-world operational settings.

References

[1] A. Khan and M. Rizwan, “Load Forecasting Using Different Techniques,” *Lecture Notes in Electrical Engineering*, vol. 956, pp. 131 – 151, 2023. https://doi.org/10.1007/978-981-19-6490-9_8

[2] R. Chandrasekaran and S. K. Paramasivan, “Advances in Deep Learning Techniques for Short-term Energy Load Forecasting Applications: A Review,” *Archives of Computational Methods in Engineering*, vol. 32, no. 2, pp. 663 – 692, 2025. https://doi.org/10.1007/s11831-024-10155-x

[3] V. Hassija *et al.*, “Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence,” Jan. 01, 2024, *Springer*. https://doi.org/10.1007/s12559-023-10179-8

[4] X. Yuan, S. Liu, W. Feng, and G. Dauphin, “Feature Importance Ranking of Random Forest-Based End-to-End Learning Algorithm,” *Remote Sens. (Basel).*, vol. 15, no. 21, Nov. 2023. https://doi.org/10.3390/rs15215203

[5] F. Cappelli, G. Castronuovo, S. Grimaldi, and V. Telesca, “Random Forest and Feature Importance Measures for Discriminating the Most Influential Environmental Factors in Predicting Cardiovascular and Respiratory Diseases,” *Int. J. Environ. Res. Public Health*, vol. 21, no. 7, Jul. 2024. https://doi.org/10.3390/ijerph21070867

[6] B. Magalhães, P. Bento, J. Pombo, M. do R. Calado, and S. Mariano, “Short-Term Load Forecasting Based on Optimized Random Forest and Optimal Feature Selection,” *Energies (Basel).*, vol. 17, no. 8, 2024. https://doi.org/10.3390/en17081926

[7] S. Zhao, “Urban Electricity Consumption Forecasting Based on SARIMA and Random Forest Modeling,” *Journal of Electrotechnology, Electrical Engineering and Management*, vol. 7, no. 1, 2024. https://doi.org/10.23977/jeeem.2024.070115

[8] B. Yang, M. Gül, and Y. Chen, “Comparative analysis of deep learning and tree-based models in power demand prediction: Accuracy, interpretability, and computational efficiency,” *J. Build. Phys.*, vol. 49, no. 1 Special Issue: eSim2024, part one, pp. 127–169, Jul. 2025. https://doi.org/10.1177/17442591251333144

[9] N. J. Johannesen, M. Kolhe, and M. Goodwin, “Relative evaluation of regression tools for urban area electrical energy demand forecasting,” *J. Clean. Prod.*, vol. 218, pp. 555 – 564, 2019. https://doi.org/10.1016/j.jclepro.2019.01.108

[10] M. L. Wallace *et al.*, “Use and misuse of random forest variable importance metrics in medicine: demonstrations through incident stroke prediction,” *BMC Med. Res. Methodol.*, vol. 23, no. 1, 2023. https://doi.org/10.1186/s12874-023-01965-x

[11] H. Prabhu, A. Sane, R. Dhadwal, N. R. Parlikkad, and J. K. Valadi, “Interpretation of Drop Size Predictions from a Random Forest Model Using Local Interpretable Model-Agnostic Explanations (LIME) in a Rotating Disc Contactor,” *Ind. Eng. Chem. Res.*, 2023. https://doi.org/10.1021/acs.iecr.3c00808

[12] R. Guidotti, “Counterfactual explanations and how to find them: literature review and benchmarking,” *Data Min. Knowl. Discov.*, vol. 38, no. 5, pp. 2770–2824, Sep. 2024. https://doi.org/10.1007/s10618-022-00831-6

[13] F. Cheng, Y. Ming, and H. Qu, “DECE: Decision Explorer with Counterfactual Explanations for Machine Learning Models,” *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 1438–1447, 2021. https://doi.org/10.1109/TVCG.2020.3030342

[14] J. Höllig, A. F. Markus, J. de Slegte, and P. Bagave, *Semantic Meaningfulness: Evaluating Counterfactual Approaches for Real-World Plausibility and Feasibility*, vol. 1902 CCIS. 2023. https://doi.org/10.1007/978-3-031-44067-0_32

[15] P. Rakesh, J. R. Shruti, I. Thippeswamy, B. L. Nithya, and V. Dheeraj, "A Surrogate Approach to Explainable AI for Predictive Maintenance: Techniques and Applications," in *5th International Conference on Circuits, Control, Communication and Computing, I4C 2024*, 2024, pp. 160 – 165. https://doi.org/10.1109/I4C62240.2024.10748524
[16] A. L. Alfeo and M. G. C. A. Cimino, "Counterfactual-Based Feature Importance for Explainable Regression of Manufacturing Production Quality Measure," in *International Conference on Pattern Recognition Applications and Methods*, 2024, pp. 48–56. https://doi.org/10.5220/0012369600003654
[17] L. Celar and R. M. J. Byrne, "How people reason with counterfactual and causal explanations for Artificial Intelligence decisions in familiar and unfamiliar domains," *Mem. Cognit.*, vol. 51, no. 7, pp. 1481–1496, 2023. https://doi.org/10.3758/s13421-023-01407-5
[18] P. M. Vannostrand, D. M. Hofmann, L. Ma, B. Genin, R. Huang, and E. A. Rundensteiner, "Counterfactual Explanation Analytics: Empowering Lay Users to Take Action Against Consequential Automated Decisions," *Proceedings of the VLDB Endowment*, vol. 17, no. 12, pp. 4349–4352, 2024. https://doi.org/10.14778/3685800.3685872
[19] Z. Wang, I. Miliou, I. Samsten, and P. Papapetrou, "Counterfactual Explanations for Time Series Forecasting," in *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2023, pp. 1391–1396. https://doi.org/10.1109/ICDM58522.2023.00180
[20] G. Zuin and A. Veloso, "Navigating Time's Possibilities: Plausible Counterfactual Explanations for Multivariate Time-Series Forecast through Genetic Algorithms," in *2024 IEEE 23rd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, IEEE, Dec. 2024, pp. 2575–2582. https://doi.org/10.1109/TrustCom63139.2024.00359
[21] D. You *et al.*, "Counterfactual explanation generation with minimal feature boundary," *Inf. Sci. (N Y).*, vol. 625, pp. 342 – 366, 2023. https://doi.org/10.1016/j.ins.2023.01.012
[22] J. Liu, X. Wu, S. Liu, and S. Gong, "Model-agnostic counterfactual explanation: A feature weights-based comprehensive causal multi-objective counterfactual framework," *Expert Syst. Appl.*, vol. 266, Mar. 2025. https://doi.org/10.1016/j.eswa.2024.126063
[23] Y. Yuan, K. McAreavey, S. Li, and W. Liu, "Multi-Granular Evaluation of Diverse Counterfactual Explanations," in *International Conference on Agents and Artificial Intelligence*, Science and Technology Publications, Lda, 2024, pp. 186–197. https://doi.org/10.5220/0012349900003636
[24] G. Dudek, "Short-term load forecasting using random forests," *Advances in Intelligent Systems and Computing*, vol. 323, pp. 821 – 828, 2015. https://doi.org/10.1007/978-3-319-11310-4_71
[25] X. Wu, J. He, P. Zhang, and J. Hu, "Power system short-term load forecasting based on improved random forest with grey relation projection," *Dianli Xitong Zidonghua/Automation of Electric Power Systems*, vol. 39, no. 12, pp. 50 – 55, 2015. https://doi.org/10.7500/AEPS20140916005
[26] A. M. Ernesto, "Short-term electricity load forecasting (Panama case study)," *Mendeley Data*, no. V1, 2021. https://doi.org/10.17632/byx7sztj59.1
[27] A. Soriano-Vargas *et al.*, "A visual analytics approach to anomaly detection in hydrocarbon reservoir time series data," *J. Pet. Sci. Eng.*, vol. 206, 2021. https://doi.org/10.1016/j.petrol.2021.108988
[28] S. M. Ghazali, N. Shaadan, and Z. Idrus, "Missing data exploration in air quality data set using r-package data visualisation tools," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 2, pp. 755 – 763, 2020. https://doi.org/10.11591/eei.v9i2.2088
[29] N. Bui, D. Nguyen, and V. A. Nguyen, "Counterfactual Plans Under Distributional Ambiguity," in *ICLR 2022 - 10th International Conference on Learning Representations*, 2022. https://doi.org/10.48550/arXiv.2201.12487
[30] M. Virgolin and S. Fracaros, "On the robustness of sparse counterfactual explanations to adverse perturbations," *Artif. Intell.*, vol. 316, 2023. https://doi.org/10.1016/j.artint.2022.103840

© 2026 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY NC SA license (https://creativecommons.org/licenses/by-nc-sa/4.0/)