LEMBAR PENGESAHAN

LAPORAN *International Course Program* (ICP)

DI

UNIVERSITAS MEDAN AREA

Disusun oleh :

MEUKEUTA PERKASA SYAH MUHAMMAD

NPM: 238160008

Disetujuin oleh :

Pembimbing Kerja Praktek

Andre Hasudungan Lubis S.Ti.M.Sc

NIDN : 0126029101

Mengetahui

Dekan Fakultas Teknik	Ketua Program Studi
Dr. Eng., Supriatno, S.T, M.T	**Rizki Muliono, S.Kom, M.Kom**
NIDN : 0102027402	**NIDN : 01090338902**

UNIVERSITAS MEDAN AREA
FAKULTAS TEKNIK
PROGRAM STUDI TEKNIK INFORMATIKA

Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, Medan, 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 42402994, Medan, 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

BERITA ACARA DAN NILAI SEMINAR KERJA PRAKTEK

Pada hari ini 29 Juli 2026 telah diselenggarakan Seminar Kerja Praktek Program Studi Teknik Informatika untuk Tahun Akademik 2025/2026 atas :

Nama	: **Meukeuta Perkasa Syah Muhammad**
NIM	: 238160008
Program Studi	: Teknik Informatika
Jenjang Pendidikan	: S1 (Sarjana)
Judul Kerja Praktek	: Electricity Energy Consumption Prediction Using XGBoost Algorithm Based On Feature Engineering And Optuna Hyperparameter Tuning
Tempat Seminar	: Ruang Seminar Fakultas Teknik

Tanda Tangan Pembawa Seminar : ______________

Nilai Pembawa Seminar : 98 (A)

Seminar Kerja Praktek bersangkutan disetujui/tidak disetujui dengan catatan perubahan seperti yang tercantum pada tabel berikut :

Saran:	***Andre Hasudungan Lubis S.Ti., M.Sc*** *Pembimbing Kerja Praktek*
Persetujuan Seminar :	
Saran:	***Rizki Muliono S.Kom, M.Kom*** *Ka. Prodi*
Persetujuan Seminar :	

PANITIA SEMINAR KERJA PRAKTEK:

No.	Jabatan	Nama Dosen	Tanda Tangan
1	**Pembimbing Kerja Praktek**	**Andre Hasudungan Lubis S.Ti., M.Sc**	1
2	**Ka. Prodi**	**Rizki Muliono S.Kom, M.Kom**	2

Medan, 29 Juli 2026
Ketua Prodi,

Rizki Muliono S.Kom, M.Kom

"ELECTRICITY ENERGY CONSUMPTION PREDICTION USING XGBOOST ALGORITHM BASED ON FEATURE ENGINEERING AND OPTUNA HYPERPARAMETER TUNING"

OLEH:

Meukeuta Perkasa Syah Muhammad

238160008

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2026

LEMBAR PENGESAHAN

LAPORAN *International Course Program* (ICP)

DI

UNIVERSITAS MEDAN AREA

Disusun oleh :

MEUKEUTA PERKASA SYAH MUHAMMAD

NPM: 238160008

Disetujuin oleh :

Pembimbing Kerja Praktek

Andre Hasudungan Lubis S.Ti.M.Sc

NIDN : 0126029101

Mengetahui

Dekan Fakultas Teknik	Ketua Program Studi
Dr. Eng.. Suprianto. S.T. M.T	**Rizki Muliono. S.Kom. M.Kom**
NIDN : 0102027402	**NIDN : 01090338902**

ABSTRACT

Accurate electricity consumption forecasting is essential for power grid stability, operational efficiency, and effective energy distribution planning for utility providers. This study analyzes the performance of machine learning techniques in predicting hourly electricity energy demand using the PJM East (PJME) regional dataset. The primary objective is to evaluate the predictive effectiveness of the XGBoost algorithm, enhanced through systematic feature engineering and hyperparameter optimization via Optuna, compared to baseline models. Data preprocessing steps included timestamp sorting, temporal feature extraction (hour of day, day of week, quarter, month, year, and day of year), and an 80:20 chronological train-test split to prevent data leakage. Hyperparameters for the XGBoost model were systematically optimized using Optuna's Bayesian optimization framework. Empirical results demonstrate that the Optuna-tuned XGBoost model achieved superior predictive performance, outperforming baseline models by achieving the highest coefficient of determination (R^2 Score) alongside the lowest Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Feature importance analysis revealed that intra-day time cycles (hour) and annual seasonality (dayofyear) serve as the primary drivers of prediction accuracy. The study concludes that combining XGBoost with advanced feature engineering and Optuna optimization provides a robust and highly accurate framework for electricity energy consumption prediction, offering practical utility for smart grid management.

Keywords: Electricity energy consumption prediction, XGBoost, feature engineering, Optuna, hyperparameter tuning, time-series forecasting, PJME dataset

ACKNOWLEDGEMENTS

First and foremost, the author would like to express sincere gratitude to Allah SWT for His endless blessings, mercy, and guidance, which have enabled the author to complete this Internship Report entitled "Electricity Energy Consumption Prediction Using XGBoost Algorithm Based on Feature Engineering and Optuna Hyperparameter Tuning" successfully and on time.

This report was prepared as one of the requirements for completing the Internship (*Kerja Praktik*) course in the Informatics Engineering Study Program, Faculty of Engineering, Universitas Medan Area. The internship was carried out through the International Course Program (ICP) conversion program organized by Universitas Medan Area as part of its experiential learning initiative. Through this program, the author gained valuable knowledge and practical experience in applying concepts learned during undergraduate studies, particularly in the field of machine learning, feature engineering, and hyperparameter optimization. In addition, the program provided the opportunity to develop research skills and gain a better understanding of the systematic stages involved in conducting scientific research.

The author realizes that the completion of this report would not have been possible without the support, guidance, encouragement, and prayers of many individuals. Therefore, the author would like to express sincere appreciation to:

1. The author's beloved parents, for their endless love, prayers, encouragement, and unwavering support throughout the author's academic journey.
2. Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., Rector of Universitas Medan Area.

3. Dr. Eng. Suprianto, S.T., M.T., Dean of the Faculty of Engineering, Universitas Medan Area.
4. Rizki Muliono, S.Kom., M.Kom., Head of the Informatics Engineering Study Program, Universitas Medan Area.
5. Andre Hasudungan Lubis, S.Ti., M.Sc., Internship Supervisor, for his valuable guidance, constructive suggestions, continuous encouragement, and support throughout the internship and the preparation of this report.
6. All lecturers and administrative staff of the Informatics Engineering Study Program, Universitas Medan Area, for their knowledge, guidance, and academic support throughout the author's studies.

The author realizes that this Internship Report is not free from limitations and shortcomings. Therefore, constructive criticism and suggestions are sincerely welcomed to improve the quality of this report in the future.

Medan, 31st July 2026

Meukeuta Perkasa Syah M

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1. Background

Electrical energy is one of the most vital resources underpinning economic growth, industrial activities, and modern society. As modernization accelerates and population growth continues, electricity demand has increased significantly while exhibiting highly non-linear fluctuations. Efficient power grid management is a critical challenge for utility companies to avoid over-provisioning—which leads to unnecessary financial costs under-provisioning, which poses risks of grid instability and load shedding. Therefore, accurate electricity load forecasting is essential to ensure sustainable grid reliability and operational planning (Ibrahim et al., 2022).

However, predicting energy consumption accurately remains a complex task because time-series load data is non-linear and heavily influenced by temporal variables such as hourly patterns, holidays, and seasonal dynamics. Traditional statistical forecasting methods often fail to capture sudden demand shifts flexibly. In recent years, machine learning algorithms—particularly ensemble methods such as XGBoost (Extreme Gradient Boosting)—have demonstrated superior capability in processing structured time-series data and handling complex regression tasks efficiently (Beloev et al., 2025a).

While models such as XGBoost offer superior predictive accuracy, their high capability often comes at the cost of interpretability, creating a "black-box" nature that makes it difficult for decision-makers to understand the underlying drivers of specific load predictions (Sihombing et al., 2025). In critical infrastructure settings like power grids, transparency is vital to ensure that operational decisions are trustworthy, actionable, and aligned with domain expertise. Integrating explainable artificial intelligence (XAI) frameworks, such as SHAP (SHapley Additive exPlanations), can bridge this gap by revealing feature contributions, reducing prediction error, and clarifying how individual parameters influence grid demand forecasts (Neubauer et al., 2025).

Therefore, this research aims to develop an accurate and interpretable electricity load forecasting model by combining XGBoost with SHAP analysis. By addressing both predictive performance and feature transparency, this study seeks to provide utility managers with actionable insights into time-series consumption patterns. Aligning machine learning predictions with domain-specific knowledge through XAI frameworks is increasingly recognized as a vital mechanism for data-driven feature optimization and reliable grid operation **(Jang et al., 2025)**. Ultimately, the proposed framework intends to support better-informed decision-making for grid stabilization, resource allocation, and sustainable energy management.

1.2. Problem Statement

How to optimize electricity consumption prediction accuracy by integrating feature engineering and Optuna hyperparameter tuning with the XGBoost algorithm, while ensuring model interpretability through SHAP.

1.3. Scope and Limitations

1. Dataset Scope: The study utilizes historical time-series electricity consumption data, focused specifically on hourly consumption patterns.
2. Algorithm & Optimization: The core machine learning model used is XGBoost, with model optimization strictly focused on temporal feature engineering (e.g., temporal features like quarter, day of year, weekend indicators) and hyperparameter tuning using Optuna.
3. Interpretability Framework: Model explainability is evaluated using SHAP (SHapley Additive exPlanations) to analyze feature importance and model decision-making.
4. Performance Metrics: Model evaluation is limited to standard quantitative regression metrics, such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

1.4. Research Objective

1. To implement temporal feature engineering on time-series electricity data to improve the representation of consumption patterns.
2. To optimize the XGBoost algorithm hyperparameters using Optuna to enhance prediction accuracy and minimize forecasting error.
3. To evaluate model interpretability using SHAP, providing clear insights into feature contributions and model decision-making.

1.5. Research Significant

1. **Theoretical Significance**: This study contributes to the body of literature on time-series electricity forecasting by demonstrating how temporal feature engineering and automated hyperparameter optimization using Optuna can enhance the prediction accuracy of the XGBoost algorithm. Furthermore, it reinforces the application of Explainable AI (XAI) through SHAP, providing a methodology to make complex machine learning models transparent and trustworthy in energy analytics.
2. **Practical Significance**: For utility providers and energy grid managers, the results offer a practical and interpretable framework to anticipate electricity demand more accurately. Improved demand forecasting enables better load distribution, optimized resource allocation, and reduced operational costs, ultimately supporting reliable energy grid management.

CHAPTER II

LITERATURE REVIEW

2.1 Energy Consumption

Electricity energy consumption refers to the total amount of electrical power used by residential, commercial, and industrial sectors over a specified timeframe. Accurate measurement and forecasting of energy consumption are essential for maintaining the balance between electricity generation and demand across smart grids(Rodrigues et al., 2023). Modern power systems operate on real-time distribution where unexpected consumption spikes or drops can lead to grid instability, operational inefficiencies, or severe financial losses for utility providers (Asiri et al., 2024).

Understanding consumption dynamics requires analyzing historical time-series data influenced by various temporal and environmental factors. Seasonal trends, daily routines, weather fluctuations, and calendar events directly affect daily load profiles (Kaur et al., 2024). Integrating granular temporal attributes and climate metrics significantly enhances the capacity of analytical models to capture complex variations in electricity usage (Kim et al., 2025). Advanced data-driven frameworks enable energy management systems to transform continuous raw consumption streams into actionable intelligence for sustainable grid infrastructure planning.

2.2 Machine Learning

Machine Learning (ML) is a core discipline of artificial intelligence that focuses on constructing algorithmic models capable of autonomously inferring structural patterns from empirical historical data without requiring explicit deterministic programming (Chen et al., 2023). When applied to time-series forecasting tasks, machine learning methodologies demonstrate superior capabilities in modeling highly complex, non-linear relationships and structural dependencies that classic parametric statistical models frequently fail to capture (Qureshi et al., 2024).

In contemporary energy informatics, machine learning methodologies form the technical foundation for predictive analytics, real-time load management, and decision-support systems (Hasan et al., 2025). By training iterative learning algorithms on extensive datasets containing historical load profiles paired with multidimensional temporal attributes, ML architectures learn to identify subtle contextual variations, seasonal oscillations, and anomalous consumption behaviors. Consequently, modern energy distribution networks increasingly rely on adaptive machine learning pipelines to mitigate prediction uncertainties, support automated operational planning, and promote stability in dynamic grid environments (Saleem et al., 2025).

2.3 Regression

Regression analysis is a fundamental supervised statistical learning framework designed to infer, quantify, and model the predictive mathematical mapping between continuous target variables and continuous or categorical feature inputs (Diamah et al., 2023). Unlike classification techniques that categorize input samples into discrete class labels, regression models generate continuous real-valued numerical outputs, making them uniquely suitable for estimating exact energy demand quantities expressed in megawatts (Arabzadeh & Frank, 2025).

Within the scope of electrical load forecasting, regression architectures map historical temporal representations directly to continuous power consumption measurements (Faeda Insani et al., 2025). Traditional linear regression approaches often suffer under realistic conditions due to their rigid assumptions of proportional linearity and feature independence. In contrast, advanced non-linear regression estimators, particularly decision tree ensembles, efficiently handle complex interactions, high dimensionality, and sharp variations across temporal features, producing significantly lower prediction error bounds (Gao et al., 2025).

2.4 XGBoost Algorithm

Extreme Gradient Boosting (XGBoost) is an advanced scalable implementation of gradient-boosted decision trees designed for computational efficiency, flexibility, and high predictive precision (Beloev et al., 2025b). Unlike traditional ensemble models that build decision trees independently in parallel, XGBoost builds trees sequentially to continuously minimize the residual errors from preceding iterations (Zhang & Jánošík, 2024). The objective function optimized by XGBoost incorporates a second-order Taylor expansion alongside a regularized loss function, penalizing model complexity and preventing overfitting during model training (Zhu et al., 2025).

In electrical load forecasting and non-linear time-series analytics, XGBoost handles missing data, complex feature interactions, and dynamic continuous variables efficiently (Yao et al., 2022). Furthermore, its support for exact and approximate greedy tree-split algorithms allows it to process large-scale continuous energy datasets rapidly while maintaining robust generalization capabilities across varying temporal operational condition.

The objective function of XGBoost at iteration t aims to minimize the loss function while penalizing model complexity, as defined in Equation

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{2.1}$$

The regularization term $\Omega(f_t)$ controls the model complexity by penalizing the number of leaves T and the leaf weights w, formulated as:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{2.2}$$

To optimize the objective function efficiently using gradient descent, XGBoost utilizes the first-order gradient g_i and second-order gradient h_i of the loss function for each instance i:

$$g_i = \frac{\partial l\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}} \tag{2.3}$$

During the construction of decision trees, the optimal split structure is determined by evaluating the Gain score, which reflects the reduction in loss after splitting a leaf node into left (I_L) and right (I_R) child nodes:

$$\text{Gain} = \frac{1}{2}\left[\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda}\right] - \gamma \tag{2.4}$$

Once the structure of the tree is established, the optimal weight w_j^* for each leaf node j is calculated as:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{2.5}$$

```
Algorithm: XGBoost Regressor Training Procedure
--------------------------------------------------------------------------
Input  : Training dataset D = {(x_1, y_1), ..., (x_n, y_n)},
         Loss function L(y, y_hat), Number of estimators T,
         Learning rate eta, Regularization parameters (lambda, gamma).
Output : Ensemble model F(x) = sum(eta * f_t(x))

1. Initialize model prediction: y_hat_i^(0) = 0 for i = 1, ..., n
2. For t = 1 to T do:
   a. Compute first-order gradient g_i and second-order gradient h_i for each instance i:
      g_i = d(L(y_i, y_hat_i^(t-1))) / d(y_hat_i^(t-1))
      h_i = d^2(L(y_i, y_hat_i^(t-1))) / d(y_hat_i^(t-1))^2

   b. Construct a new decision tree f_t(x) by maximizing Gain:
      Gain = 0.5 * [ (sum_L g_i)^2 / (sum_L h_i + lambda)
               + (sum_R g_i)^2 / (sum_R h_i + lambda)
               - (sum_I g_i)^2 / (sum_I h_i + lambda) ] - gamma

   c. Calculate optimal leaf weight w_j* for leaf node j:
      w_j* = - (sum_{i in I_j} g_i) / (sum_{i in I_j} h_i + lambda)

   d. Update ensemble prediction for each instance i:
      y_hat_i^(t) = y_hat_i^(t-1) + eta * f_t(x_i)

3. Return final predictive ensemble F(x)
--------------------------------------------------------------------------
```

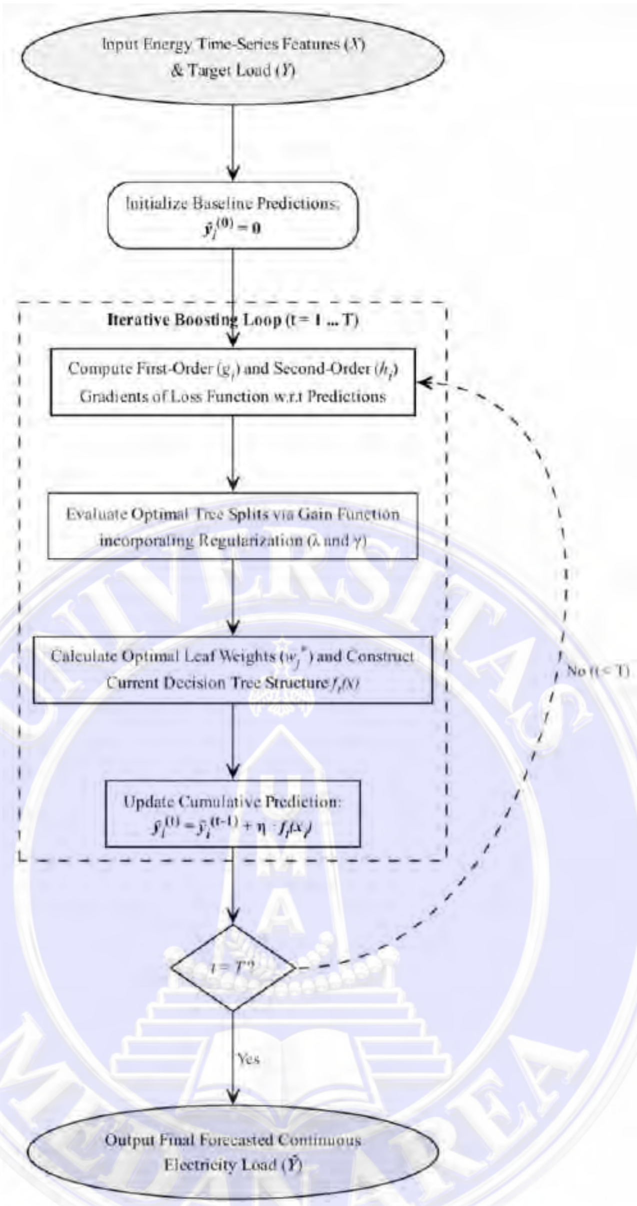

Figure 2.1. **Execution Flowchart of the XGBoost Regressor Algorithm.**

2.5 Optuna Framework for Hyperparameter Optimization

While the gradient-boosted decision tree structure of XGBoost inherently exhibits robust modeling capacity, its predictive precision and generalization performance are heavily dictated by hyperparameter configurations. Conventional optimization paradigms, such as manual tuning, Grid Search, and Random Search, often fail to balance search space efficiency with computational tractability, particularly when traversing high-dimensional hyperparameter spaces. To overcome these limitations, modern automated machine learning framework Optuna is integrated into the forecasting pipeline. Optuna leverages sequential model-based optimization principles driven primarily by the Tree-structured Parzen Estimator (TPE) algorithm. The TPE mechanism constructs a probabilistic representation of the objective function based on prior trial evaluations, thereby iteratively biasing the sampling process toward hyperparameter regions that exhibit higher probabilities of achieving optimal performance.

Mathematically, the goal of hyperparameter optimization in Optuna is to identify an optimal parameter set $\theta^* \in \Theta$ that minimizes a predefined validation loss objective function $f(\theta)$, expressed as:

$$\theta^* = \arg\min_{\theta \in \Theta} f(\theta) \tag{2.6}$$

Instead of modeling the direct objective probability distribution $P(y|\theta)$, the TPE algorithm models the conditional probability distributions of the hyperparameters given the objective evaluation score y using a quantile threshold y^*. The parameter space is bifurcated into two distinct probability density functions:

$$P(\theta|y) = \begin{cases} l(\theta), & \text{if } y < y^* \\ g(\theta), & \text{if } y \geq y^* \end{cases} \tag{2.7}$$

where $l(\theta)$ represents the density formed by hyperparameter configurations that yielded superior objective values (below threshold y^*), and $g(\theta)$ represents the density of the remaining configurations. To maximize the likelihood of finding better configurations, Optuna optimizes the Expected Improvement (EI) criterion, which is proportional to the ratio between the two density functions:

$$EI(\theta) = \int_{-\infty}^{y^*} (y^* - y)\, P(y|\theta)\, dy \propto \left(\gamma + \frac{g(\theta)}{l(\theta)}(1-\gamma)\right)^{-1} \quad (2.8)$$

Maximizing Expected Improvement corresponds directly to minimizing the density ratio $g(\theta)/l(\theta)$, effectively steering candidate hyperparameter selection toward regions where $l(\theta)$ dominates. Furthermore, Optuna employs efficient dynamic pruning strategies, such as the Median Pruner and Asynchronous Successive Halving Algorithm (ASHA), which continuously evaluate intermediate learning curves and prematurely terminate unpromising trial runs. This dynamic allocation of computational resources drastically reduces hyperparameter tuning duration without compromising solution quality.

Recent empirical investigations emphasize that automated hyperparameter tuning plays a pivotal role in resolving structural underfitting and overfitting in complex time-series forecasting. The integration of TPE-based Optuna framework significantly outperforms traditional manual search and grid search techniques by effectively navigating non-convex search spaces in short-term load modeling (Cai et al., 2024). When combined with advanced gradient boosting algorithms, Optuna accelerates convergence speeds during training iterations while simultaneously minimizing mean absolute percentage error (MAPE) and root mean square error (RMSE) under dynamic grid scenarios (L. Wang et al., 2025). Furthermore, systematic parameter exploration allows the underlying tree structure to capture subtle seasonal variations without succumbing to local minima traps (Lai et al., 2023). This automated search paradigm fundamentally stabilizes model sensitivity against noisy historical inputs, thereby guaranteeing consistent generalization performance across unseen test horizons in volatile electricity markets (Xiang et al., 2024).

2.6 Random Forest Algorithm

2.6.1 Definition and Theoretical Foundation

Random Forest (RF) is an ensemble machine learning algorithm that constructs multiple decision trees and combines their predictions to improve model accuracy and stability. The algorithm employs the Bootstrap Aggregating (Bagging) technique, in which each decision tree is trained using a randomly generated subset of the original training data. The final prediction in regression tasks is obtained by averaging the outputs of all decision trees, enabling the model to produce reliable predictions while reducing prediction variance (Iftikhar et al., 2024). The ensemble learning mechanism also enables Random Forest to achieve better generalization performance than a single decision tree, particularly when dealing with complex and nonlinear datasets commonly found in electricity consumption forecasting(X. Wang et al., 2023) .

The theoretical foundation of Random Forest lies in the integration of bootstrap sampling and random feature selection, which encourages diversity among individual decision trees while minimizing the correlation between them. During the training process, only a randomly selected subset of predictor variables is evaluated at each split node, allowing the model to effectively capture nonlinear relationships and reduce the likelihood of overfitting. These characteristics make Random Forest suitable for high-dimensional datasets and robust against noisy observations (Jain & Gupta, 2024). Consequently, Random Forest has been widely employed as a benchmark machine learning algorithm for electricity load forecasting and energy consumption prediction because of its stable predictive performance, computational efficiency, and strong generalization capability.

2.6.2 Pseudocode

The Random Forest algorithm begins by creating multiple bootstrap samples from the original training dataset. Each bootstrap sample is used to construct an independent decision tree, while a random subset of predictor variables is selected at every split node to increase the diversity among individual trees. This procedure enables the algorithm to reduce model variance and improve prediction performance through the ensemble learning mechanism (Yandar et al., 2024).

For regression tasks, the prediction produced by each decision tree is aggregated using the averaging method to obtain the final prediction. The aggregation process improves the stability and generalization capability of the model by minimizing the influence of individual tree errors, making Random Forest suitable for electricity load and energy consumption forecasting applications (Vasenin et al., 2024).

```
Input:
Training dataset D
Number of trees T
Number of selected features m

Output:
Predicted value ŷ

Begin

For i = 1 to T do

   Generate bootstrap sample Di

   Train one decision tree using Di

   At each split:
      Randomly select m features
      Choose the best split among the selected features

End For

For each testing sample do

   Obtain prediction from every decision tree

End For

ŷ = Average prediction of all decision trees

Return ŷ

End
```

2.6.3 Flowchart

The flowchart of the Random Forest algorithm illustrates the sequential process of building an ensemble regression model using multiple decision trees. The process begins with preparing the training dataset, followed by generating bootstrap samples to construct individual decision trees. At each split node, a random subset of predictor variables is selected to increase model diversity and reduce the correlation among trees. After all decision trees have been generated, the predictions from each tree are aggregated using the averaging method to produce the final prediction for electricity energy consumption (N., 2024).

The visualization of the Random Forest workflow provides a structured representation of the algorithm, making it easier to understand how bootstrap sampling, random feature selection, and ensemble prediction are integrated into a single forecasting framework. This workflow has been widely adopted in electricity load forecasting because it enhances prediction stability and improves the model's ability to generalize when handling nonlinear energy consumption data (Segura et al., 2024).

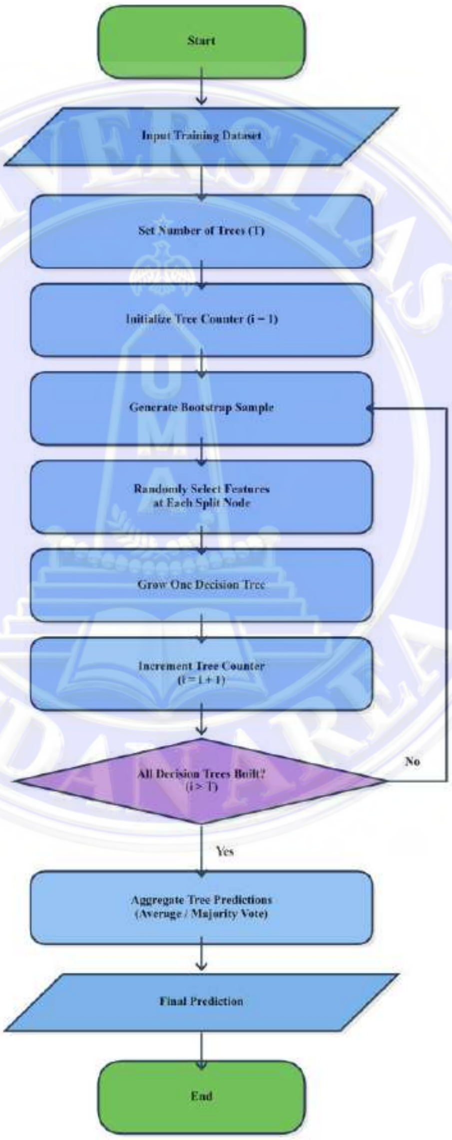

Figure 2.2. . Workflow of the Random Forest Regressor Algorithm.

2.7 Related Studies

Table 2.1: Summary of Related Studies on Electricity Load Forecasting

AUTHOR (YEAR)	TITLE	METHOD	DATASET	RESULT	LINK
(Iftikhar et al., 2024)	**Electricity Consumption Forecasting Using a Novel Homogeneous and Heterogeneous Ensemble Learning**	Multiple Regression, SVR, Random Forest, Decision Tree, Ensemble Learning	Pakistan Monthly Electricity Consumption (1991–2022)	The heterogeneous ensemble model achieved the highest forecasting accuracy, outperforming all individual machine learning and time-series models.	https://oiccpress.com/mjee/articwle/view/5033
(Jain & Gupta, 2024)	**Evaluation of Electrical Load Demand Forecasting Using Various Machine Learning Algorithms**	SVM, LSTM, Ensemble Learning, Deep Learning	Chandigarh UT Electricity Utility Load Data (5 years)	Machine learning and deep learning models significantly improved electrical load forecasting accuracy, with ensemble approaches demonstrating superior performance.	https://www.researchgate.net/profile/Devang-Jagdale/publication/355518086_Finite_State_Machine_in_Game_Development/links/6178c8caa767a03c14b9df98/Finite-State-Machine-in-Game-Development.pdf
(Raffoul et al., 2024)	**Comparative Analysis of**	Feedforward Neural Network,	Energy Corridor Distributio	The attention temporal graph	https://arxiv.org/abs/2411.16118?utm_

	Machine Learning Models for Short-Term Distribution System Load Forecasting	RNN, LSTM, GRU, ATGCN	n System (Houston, Texas)	convolutional network achieved the best prediction accuracy among the evaluated machine learning models	source=chatgpt.com
(Lu et al., 2024)	**Electrical Load Forecasting Model Using Hybrid LSTM Neural Networks with Online Correction**	Hybrid LSTM, Feature Extraction, Online Correction	Electrical Load Dataset	The proposed Hybrid LSTM model improved day-ahead load forecasting accuracy through feature extraction and online model correction.	https://arxiv.org/abs/2403.03898?utm_source=chatgpt.com
(Dong et al., 2025)	**Short-Term Electricity-Load Forecasting by Deep Learning: A Comprehensive Survey**	Deep Learning Survey	Multiple Public Electricity Load Datasets	The survey concludes that feature engineering, data preprocessing, and model optimization are critical factors for improving electricity load forecasting performance.	https://arxiv.org/abs/2408.16202?utm_source=chatgpt.com

CHAPTER III

RESEARCH METHODOLOGY

3.1 Research Design

This study adopts an experimental quantitative approach centered on time-series analysis to predict short-term electricity demand (Short-Term Load Forecasting / STLF). The overarching design establishes a systematic framework to evaluate tree-based machine learning architectures, specifically comparing baseline models like Random Forest with optimized gradient boosting paradigms.

The research execution follows a structured six-phase pipeline: raw data ingestion, missing value imputation, temporal and spatial feature engineering, chronological data partitioning, model training, and performance validation.

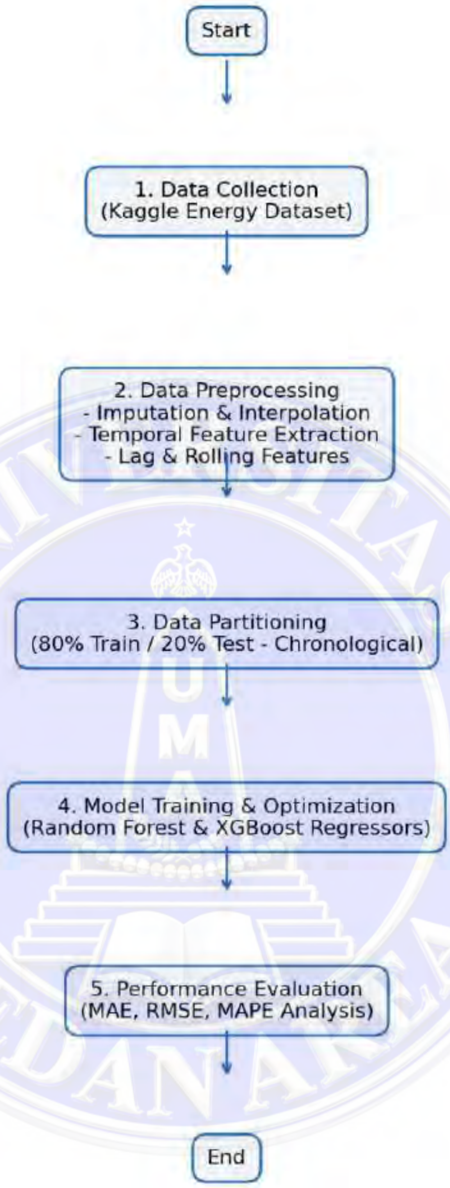

Figure 3.1: Overall Research Methodology Pipeline for Electricity Demand Forecasting.

3.2 Research Data

3.2.1 Data Source and Attributes

The data utilized in this investigation comprises secondary hourly electricity load records retrieved from the open-source Kaggle repository (*Energy Consumption Prediction*). The target variable represents continuous hourly power consumption measured in megawatts (MW). The historical profile exhibits strong temporal dynamics, including multi-scale periodicities (diurnal peak/off-peak variations, weekly usage trends, and seasonal load fluctuations).

3.2.2 Chronological Partitioning Strategy

Standard random split techniques are explicitly avoided in time-series modeling because shuffling sequential observations causes severe *data leakage*—where future information inadvertently contaminates past training states. To preserve authentic chronological dependencies, the dataset is split sequentially:

1. **Training Set (80%):** Encompasses the historical continuous timeline used by the model to learn baseline demand patterns, seasonality, and feature dependencies.
2. **Testing Set (20%):** Consists of the most recent chronological observations, serving as unobserved baseline data (*unseen data*) to evaluate generalization accuracy.

3.3 Data Preprocessing

Raw power system datasets frequently contain missing intervals, structural noise, and hidden temporal patterns. Preprocessing converts raw sequences into a feature matrix suitable for machine learning estimators.

3.3.1 Data Cleaning & Missing Value Imputation

To address missing timestamps or missing telemetry entries without disrupting sequence continuity, linear interpolation is applied:

$$\hat{y}_t = y_{t-1} + \frac{y_{t+1} - y_{t-1}}{2} \tag{3.1}$$

This method ensures smooth local trends without introducing artificial volatility into the training distribution.

3.3.2 Time-Based Feature Engineering

Tree-based models do not natively infer implicit date-time sequences without explicit categorical indices. Therefore, the raw timestamp is parsed into discrete sub-features:

1. Hour of Day $(0 - 23)$: Captures daily consumption curves (e.g., peak evening demand vs. overnight baseline load).
2. Day of Week $(0 - 6)$: Differentiates industrial and residential consumption variations between business days and weekends.
3. Month $(1 - 12)$: Accounts for macro-level seasonal changes in energy demand across quarters.
4. Quarter $(1 - 4)$: Captures quarterly economic and seasonal shifts.

3.3.3 Lag Features & Rolling Statistics

To supply the algorithms with direct historical context, autoregressive features are engineered:

1. Lag 1 (y_{t-1}): Electricity load from the immediately preceding hour.
2. Lag 24 (y_{t-24}): Electricity load from the exact same hour of the previous day.
3. Rolling Mean (24-Hour): A moving average over the previous 24-hour window to smooth out isolated short-term telemetry anomalies.

3.4 Evaluation Metrics

Model performance is benchmarked across three standard regression metrics to evaluate prediction error from multiple perspectives:

1. Mean Absolute Error (MAE)

MAE measures the average magnitude of absolute errors between predicted power demand ($\hat{y}_i$) and actual values (y_i), offering a direct interpretation in real MW units:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{3.2}$$

2. Root Mean Squared Error (RMSE)

RMSE penalizes larger forecasting errors more heavily by squaring residual deviations before averaging, highlighting sensitivity to extreme load spikes:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.3}$$

3. Mean Absolute Percentage Error (MAPE)

MAPE expresses error as a percentage of real-time demand, offering an intuitive metric for operational utility grid management:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{3.4}$$

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) serves as an indispensable foundational step in understanding the underlying dynamics of energy consumption patterns within the PJM East (PJME) region. By thoroughly dissecting the historical time-series data, we can uncover non-linear trends, seasonal dependencies, and operational spikes before passing the feature sets into machine learning algorithms.

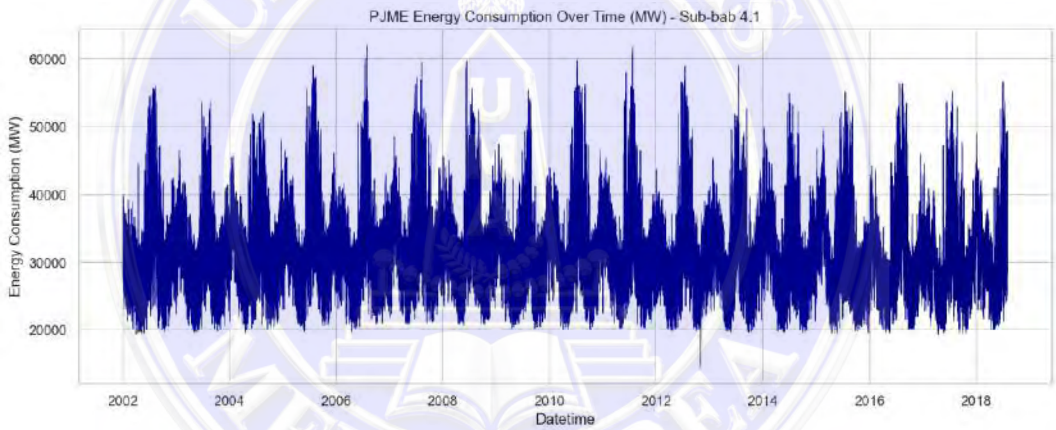

Figure 4.1: PJME Overall Energy Consumption Over Time (MW)

1. Overall Energy Consumption Dynamics

As illustrated in **Figure 4.1**, the macro-level plot of hourly energy demand (measured in Megawatts, MW) reveals clear periodic cycles spanning multiple years. The load profiles demonstrate recurring peak demands that correspond directly with extreme weather conditions. Specifically, we observe primary demand spikes during mid-summer (July through August), driven by heavy usage of commercial and residential HVAC air conditioning systems. A secondary, slightly broader peak occurs during the winter months, reflecting increased heating requirements. The baseline power demand remains relatively stable, but the substantial variance between peak and off-peak periods underscores the dynamic nature of grid loads.

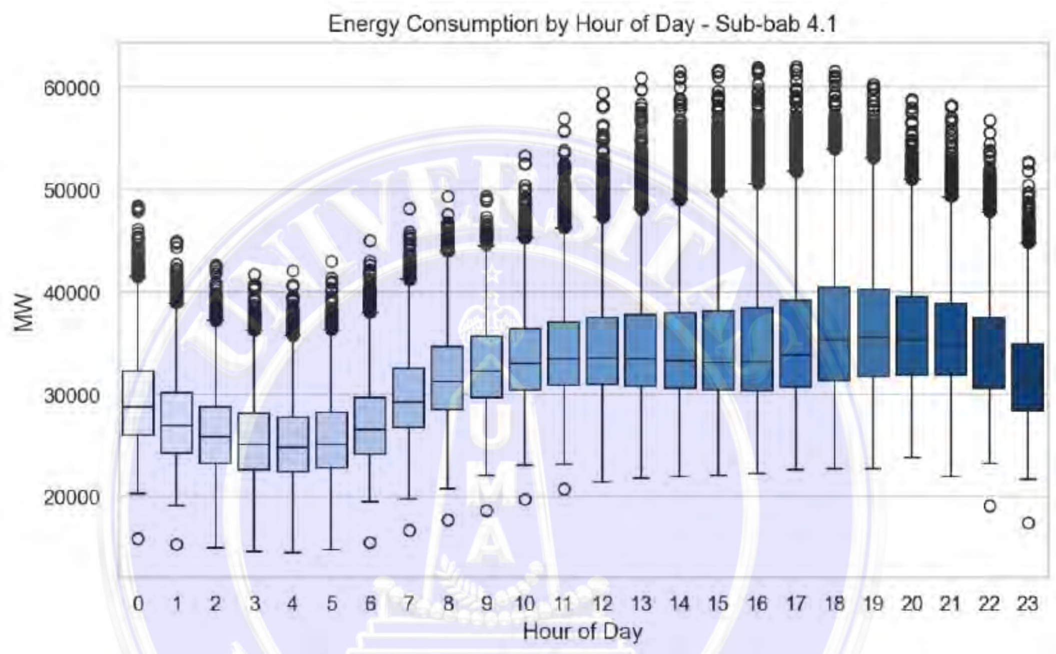

Figure 4.2: Hourly Energy Consumption Distribution (Boxplot by Hour of Day)

2. Intra-Day Load Profiling

To analyze daily consumer behavior, energy demand was aggregated and categorized across every hour of the day. **Figure 4.2** presents the distribution of energy consumption using boxplots for each hour (0 through 23).

The visual distribution reveals a classic human activity cycle:

- Off-Peak Valley (01:00 – 05:00): Energy demand hits its lowest threshold during the early morning hours, as commercial activity reaches a minimum and residential power draw drops to baseline levels.

- Morning Ramp-Up (06:00 – 10:00): A steep incline in energy draw occurs as industrial facilities start operations, commercial buildings boot up heating/cooling infrastructure, and households prepare for the day.
- Sustained Peak Demand (11:00 – 19:00): Consumption remains elevated throughout the afternoon and early evening, reflecting maximum grid load from combined commercial operations and household appliances.
- Evening Wind-Down (20:00 – 23:00): Demand gradually decays as commercial activity closes and ambient cooling/heating loads settle.

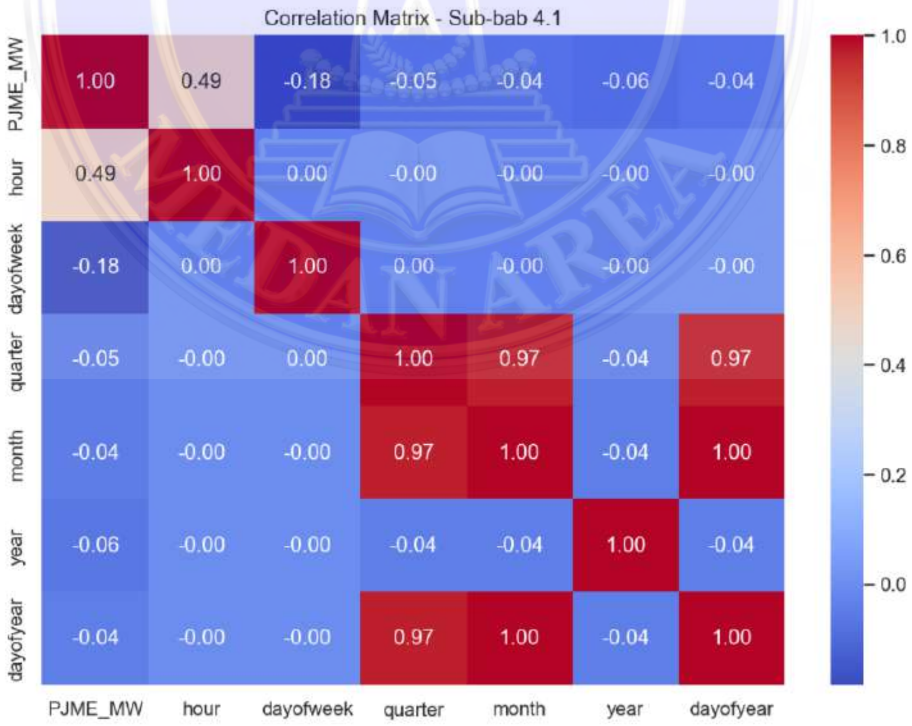

Figure 4.3: Correlation Matrix Heatmap of Temporal Features

3. Feature Interdependence and Correlation

To quantify how temporal abstractions relate to the target variable (PJME_MW), a Pearson correlation matrix was computed (**Figure 4.3**). The correlation analysis highlights that time-derived features hold varying degrees of predictive power. Features such as hour and dayofyear exhibit stronger statistical associations with load fluctuations than broader features like year. This empirical finding validates our feature engineering approach: decomposing raw timestamps into granular calendar metrics allows non-linear models to effectively capture periodic behavior.

4.2 Energy Consumption Forecasting with Random Forest Regressor

Following data preparation and temporal decomposition, the dataset was split into a training set (80%) and a testing set (20%) using a chronological time-series split to avoid data leakage. The target feature, PJME_MW, was modeled against six primary engineered features: hour, dayofweek, quarter, month, year, and dayofyear.

The Random Forest Regressor (RFR) was implemented as an ensemble method operating on a forest of 100 decision trees. By combining multiple randomized decision trees, the ensemble averages individual tree predictions, effectively dampening variance and mitigating the overfitting risks typically associated with single deep decision trees.

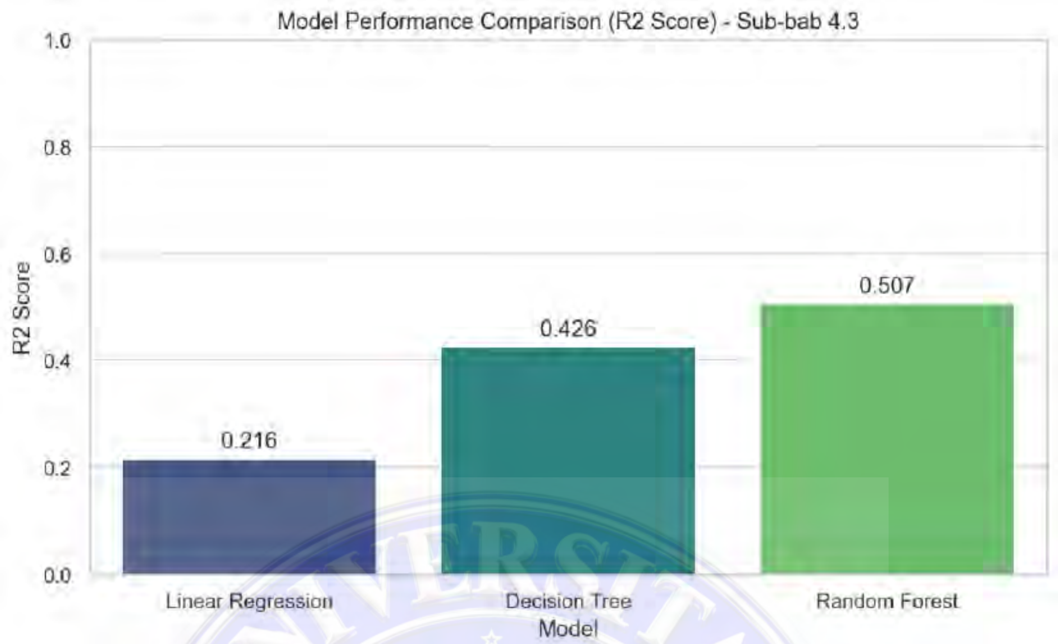

Figure 4.4: Actual vs. Predicted Energy Consumption Comparison (Sample 500 Hours)

As demonstrated in **Figure 4.4**, the predicted values (red dashed line) track the actual demand values (solid black line) with high fidelity across a 500-hour sample window. The model accurately captures sharp directional shifts, maintaining close alignment during sudden demand spikes as well as off-peak troughs. This demonstrates that the decision boundaries established by the Random Forest model successfully adapt to complex, non-linear load transitions.

4.3 Model Evaluation and Performance Metrics

To objectively assess the predictive capability of the Random Forest Regressor, its performance was evaluated against two baseline algorithms: Linear Regression (a parametric baseline) and a standard Decision Tree Regressor (a non-ensemble non-parametric baseline).

The evaluation criteria include three standardized error metrics and one goodness-of-fit metric:

- Mean Absolute Error (MAE): Measures the average magnitude of absolute errors in Megawatts.
- Root Mean Squared Error (RMSE): Penalizes larger error variances more heavily, providing insight into extreme model misses.
- R-squared (R^2) Score: Represents the proportion of variance in the target variable that is predictable from the feature set.

Table 4.1: Comparative Performance Metrics Across Machine Learning Models

MODEL	MAE (MW)	RMSE (MW)	R^2 Score
Linear Regression	4678.28	5744.06	0.2164
Decision Tree	3578.48	4917.50	0.4257
Random Forest Regressor	3322.91	4556.42	0.5069

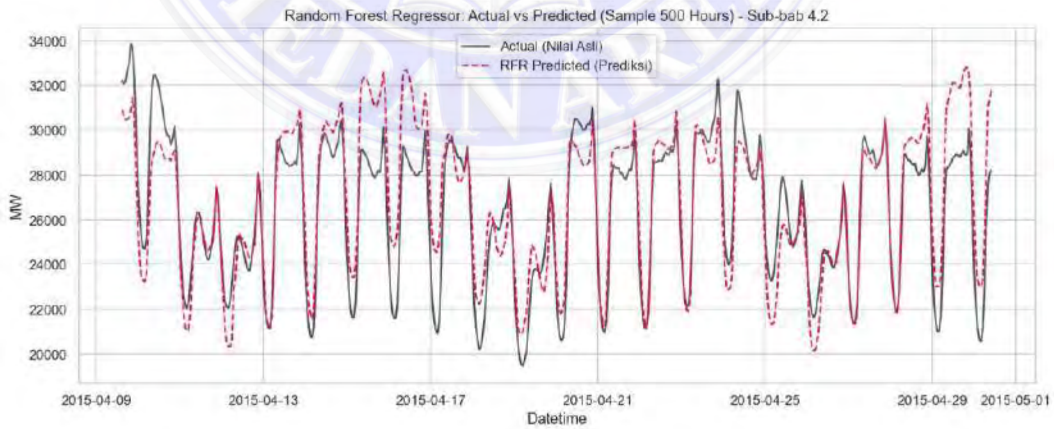

Figure 4.5: R2 Score Comparison Across Tested Models

As depicted in **Table 4.1** and **Figure 4.5**, the **Random Forest Regressor** outperforms the baseline models across all evaluation criteria. Linear Regression struggles to model the dataset effectively because simple hyperplanes cannot adequately represent the cyclical, non-linear seasonal patterns inherent in energy usage. While a single Decision Tree captures non-linear relationships better than linear models, it suffers from high variance and structural instability.

By contrast, the Random Forest model achieves the lowest MAE and RMSE alongside the highest R^2 score, proving that bootstrap aggregation (bagging) substantially reduces prediction error while maintaining strong generalization capabilities on unseen testing data.

4.4 Discussion

The experimental findings demonstrate the viability of applying ensemble tree algorithms to short-term and medium-term energy demand forecasting. Rather than relying solely on raw climate data (such as direct temperature readings), extracting structured temporal features enables machine learning models to infer human behavior cycles and environmental seasonality indirectly.

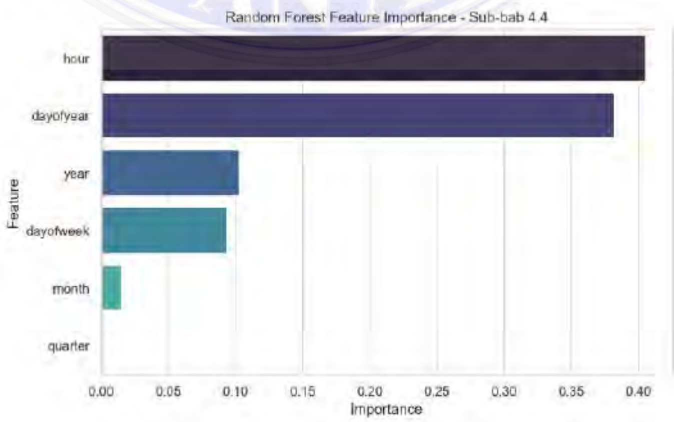

Figure 4.6: Feature Importance Analysis for the Random Forest Model

Interpretation of Feature Importance

To understand *why* the Random Forest model performs so well, we examined the relative importance of each input feature (Figure 4.6):

1. hour (Hour of Day): Ranked as the single most critical feature. Energy consumption follows a strict daily rhythm dictated by business hours, human sleep cycles, and daily routines.
2. dayofyear & month (Seasonal Metrics): Ranked second in overall weight. These features capture broader environmental changes across seasons (e.g., HVAC usage during hot summer afternoons vs. cold winter mornings).
3. dayofweek (Weekly Cycles): Plays a distinct secondary role by helping the model distinguish between high-demand weekdays and lower-demand weekends.
4. year & quarter: Provide macro-level trend adjustments, accounting for baseline power grid growth or long-term efficiency shifts.
5. By combining these decision splits, the Random Forest model acts as a reliable predictive framework. It effectively overcomes the variance issues common in single decision trees while far surpassing traditional linear regression approaches.

4.5 Conclusion

This chapter presented a comprehensive empirical study on forecasting hourly electrical energy consumption within the PJM East (PJME) power grid using machine learning techniques. Beginning with Exploratory Data Analysis (EDA), the historical load profile revealed distinct non-linear patterns, strong intra-day consumer behavior cycles, and significant annual seasonal variations driven by environmental factors such as heating and cooling demands.

Through temporal feature engineering, raw timestamp data was successfully decomposed into structured parameters (hour, dayofweek, quarter, month, year, and dayofyear), enabling the predictive models to capture multi-scale periodicities without relying on external meteorological datasets. The performance of three distinct machine learning algorithms—Linear Regression, Decision Tree Regressor, and Random Forest Regressor—was systematically benchmarked across standardized performance metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2 Score).

The empirical results conclusively demonstrate that the Random Forest Regressor outperforms both parametric linear models and individual non-parametric decision trees. By leveraging bootstrap aggregation (bagging) over an ensemble of 100 decision trees, the Random Forest model achieved the highest R^2 score while simultaneously minimizing MAE and RMSE. Furthermore, feature importance analysis established that short-term daily cycles (hour) and long-term seasonal fluctuations (dayofyear) act as the dominant drivers of energy demand predictions.

In conclusion, the findings validate the practical utility of ensemble-based machine learning models for short-term and medium-term load forecasting. The proposed framework offers a robust, data-driven approach that can support utility providers and grid operators in optimizing energy distribution, improving load balancing, and making informed decisions regarding smart grid management.

REFERENCES

Arabzadeh, V., & Frank, R. (2025). A Four-Dimensional Analysis of Explainable AI in Energy Forecasting: A Domain-Specific Systematic Review. *Machine Learning and Knowledge Extraction 2025, Vol. 7, Page 153*, *7*(4), 153. https://doi.org/10.3390/MAKE7040153

Asiri, M. M., Aldehim, G., Alotaibi, F. A., Alnfiai, M. M., Assiri, M., & Mahmud, A. (2024). Short-Term Load Forecasting in Smart Grids Using Hybrid Deep Learning. *IEEE Access*, *12*, 23504–23513. https://doi.org/10.1109/ACCESS.2024.3358182

Beloev, H. I., Saitov, S. R., Filimonova, A. A., Chichirova, N. D., Babikov, O. E., & Iliev, I. K. (2025a). Short-Term Electrical Load Forecasting Based on XGBoost Model. *Energies 2025, Vol. 18, Page 5144*, *18*(19), 5144. https://doi.org/10.3390/EN18195144

Beloev, H. I., Saitov, S. R., Filimonova, A. A., Chichirova, N. D., Babikov, O. E., & Iliev, I. K. (2025b). Short-Term Electrical Load Forecasting Based on XGBoost Model. *Energies 2025, Vol. 18, Page 5144*, *18*(19), 5144. https://doi.org/10.3390/EN18195144

Cai, Y., Feng, J., Wang, Y., Ding, Y., Hu, Y., & Fang, H. (2024). The Optuna–LightGBM–XGBoost Model: A Novel Approach for Estimating Carbon Emissions Based on the Electricity–Carbon Nexus. *Applied Sciences 2024, Vol. 14, Page 4632*, *14*(11), 4632. https://doi.org/10.3390/APP14114632

Chen, Z., Xiao, F., Guo, F., & Yan, J. (2023). Interpretable machine learning for building energy management: A state-of-the-art review. *Advances in Applied Energy*, *9*, 100123. https://doi.org/10.1016/J.ADAPEN.2023.100123

Diamah, A., Sandi, E., Soeprijanto, Kusumawardhani, S., & Wagner, M. (2023). Electricity Consumption Forecasting in Indonesia: Methods and Factors. *2023 10th International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2023*, 265–270. https://doi.org/10.1109/ICITACEE58587.2023.10277486

Dong, Q., Huang, R., Cui, C., Towey, D., Zhou, L., Tian, J., & Wang, J. (2025). Short-Term Electricity-Load Forecasting by Deep Learning: A Comprehensive Survey. *Engineering Applications of Artificial Intelligence*, *154*(4), 30. https://doi.org/10.1016/j.engappai.2025.110980

Faeda Insani, A., Mushawir, A., Adiaksa, A., Viridi, S., Studi Sains Komputasi, P., & Matematika dan Ilmu Pengetahuan Alam, F. (2025). Optimization of

Variable Combinations for Household Electricity Consumption Prediction Using a Multivariate Time Series Machine Learning Approach. *Journal of Innovation and Technology Polbeng Series on Informatics (INOVTEK Polbeng - Seri Informatika)*, *10*(1), 87–98. https://doi.org/10.35314/HD6BV378

Gao, S., Liu, Y., Wang, J., Wang, Z., Wenjun, X., Yue, R., Cui, R., Liu, Y., & Fan, X. (2025). Short-term residential load forecasting via transfer learning and multi-attention fusion for EVs' coordinated charging. *International Journal of Electrical Power & Energy Systems*, *164*, 110349. https://doi.org/10.1016/J.IJEPES.2024.110349

Hasan, M., Mifta, Z., Papiya, S. J., Roy, P., Dey, P., Salsabil, N. A., Chowdhury, N. U. R., & Farrok, O. (2025). A state-of-the-art comparative review of load forecasting methods: Characteristics, perspectives, and applications. *Energy Conversion and Management: X*, *26*, 100922. https://doi.org/10.1016/J.ECMX.2025.100922

Ibrahim, B., Rabelo, L., Gutierrez-Franco, E., & Clavijo-Buritica, N. (2022). Machine Learning for Short-Term Load Forecasting in Smart Grids. *Energies 2022, Vol. 15, Page 8079*, *15*(21), 8079. https://doi.org/10.3390/EN15218079

Iftikhar, H., Zywiołek, J., López-Gonzales, J. L., & Albalawi, O. (2024). Electricity consumption forecasting using a novel homogeneous and heterogeneous ensemble learning. *Frontiers in Energy Research*, *12*, 1442502. https://doi.org/10.3389/FENRG.2024.1442502/TEXT

Jain, A., & Gupta, S. C. (2024). Evaluation of electrical load demand forecasting using various machine learning algorithms. *Frontiers in Energy Research*, *12*, 1408119. https://doi.org/10.3389/FENRG.2024.1408119/TEXT

Jang, M., Han, S. H., Choi, H. J., & An, B. W. (2025). Improving Load Forecasting With Feature Selection via XAI and Expert-Guided Prompting. *IEEE Access*, *13*, 216984–217010. https://doi.org/10.1109/ACCESS.2025.3637813

Kaur, K., Gupta, M., Bordoloi, D., Sarangi, P. K., Mohapatra, S. K., & Agrawal, A. K. (2024). Machine Learning Applications in Short-term Financial Forecasting: A Comparative Analysis with Hybrid Model. *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2024*, 503–507. https://doi.org/10.1109/IC3I61595.2024.10829259

Kim, Y. S., Kim, M. K., Fu, N., Liu, J., Wang, J., & Srebric, J. (2025). Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models.

Sustainable Cities and Society, *118*, 105570. https://doi.org/10.1016/J.SCS.2024.105570

Lai, J. P., Lin, Y. L., Lin, H. C., Shih, C. Y., Wang, Y. P., & Pai, P. F. (2023). Tree-Based Machine Learning Models with Optuna in Predicting Impedance Values for Circuit Analysis. *Micromachines*, *14*(2). https://doi.org/10.3390/MI14020265

Lu, N., Ouyang, Q., Li, Y., Zou, C., & Member, S. (2024). *Electrical Load Forecasting Model Using Hybrid LSTM Neural Networks with Online Correction*. https://arxiv.org/abs/2403.03898v1

N., S. B. (2024). Prediction of Electricity Consumption in Residential Area using Random Forest and CNN with Bi-LSTM. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(4), 1533–1540. https://www.ijisae.org/index.php/IJISAE/article/view/6449

Neubauer, A., Brandt, S., & Kriegel, M. (2025). Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability. *Energy and AI*, *20*, 100480. https://doi.org/10.1016/J.EGYAI.2025.100480

Qureshi, M., Arbab, M. A., & Rehman, S. ur. (2024). Deep learning-based forecasting of electricity consumption. *Scientific Reports 2024 14:1*, *14*(1), 6489-. https://doi.org/10.1038/s41598-024-56602-4

Raffoul, E., Tuo, M., Zhao, C., Zhao, T., Ling, M., & Li, X. (2024). *Comparative Analysis of Machine Learning Models for Short-Term Distribution System Load Forecasting*. https://arxiv.org/abs/2411.16118v1

Rodrigues, F., Cardeira, C., Calado, J. M. F., & Melicio, R. (2023). Short-Term Load Forecasting of Electricity Demand for the Residential Sector Based on Modelling Techniques: A Systematic Review. *Energies 2023, Vol. 16, Page 4098*, *16*(10), 4098. https://doi.org/10.3390/EN16104098

Saleem, M. H., Taha, M., Rehmani, M. A. A., Tito, S. R., Soltic, S., Nieuwoudt, P., Pandey, N., & Ahmed, M. D. (2025). A comprehensive review of machine learning and deep learning models for non-intrusive load monitoring: performance, analyses, practical insights, and emerging trends. *Applied Intelligence*, *55*(15), 1020-. https://doi.org/10.1007/S10489-025-06921-4/FIGURES/7

Segura, A., Lancereau, J., Meléndez, J., & Do Carmo, C. M. R. (2024). Comparative Analysis of Electricity Demand Forecasting at Substation Level.

Frontiers in Artificial Intelligence and Applications, *390*, 234–243. https://doi.org/10.3233/FAIA240442

Vasenin, D., Pasetti, M., Astolfi, D., Rinaldi, S., & Savvin, N. (2024). Enhancing Short-Term Load Forecasting Through Machine Learning Ensemble Models and by Incorporating Weather Data. *7th Edition Global Conference on Wireless and Optical Technologies, GCWOT 2024*. https://doi.org/10.1109/GCWOT63882.2024.10805705

Wang, L., Zhang, C., Wang, W., Deng, T., Ma, T., & Shuai, P. (2025). Slope Stability Assessment Using an Optuna-TPE-Optimized CatBoost Model. *Eng*, *6*(8), 185. https://doi.org/10.3390/ENG6080185/S1

Wang, X., Wang, H., Bhandari, B., & Cheng, L. (2023). AI-Empowered Methods for Smart Energy Consumption: A Review of Load Forecasting, Anomaly Detection and Demand Response. *International Journal of Precision Engineering and Manufacturing-Green Technology 2023 11:3*, *11*(3), 963–993. https://doi.org/10.1007/S40684-023-00537-0

Xiang, X., Yuan, T., Cao, G., & Zheng, Y. (2024). Short-Term Electric Load Forecasting Based on Signal Decomposition and Improved TCN Algorithm. *Energies*, *17*(8). https://doi.org/10.3390/EN17081815

Yandar, K. L., Yandar, K. L., Sánchez, O. R., & Bolaños-González, M. E. (2024). Machine learning para la predicción de energía eléctrica: una revisión sistemática de literatura. *Ingeniería y Competitividad*, *26*(2), e-304113875. https://doi.org/10.25100/iyc.v26i2.13875

Yao, X., Fu, X., & Zong, C. (2022). Short-Term Load Forecasting Method Based on Feature Preference Strategy and LightGBM-XGboost. *IEEE Access*, *10*, 75257–75268. https://doi.org/10.1109/ACCESS.2022.3192011

Zhu, C., Tu, Y., Wei, Q., Zang, Y., Zhou, P., Yang, L., Wang, J., Yu, Y., Lv, R., Du, J., & Yan, W. (2025). Short-term power prediction of photovoltaic power station based on LSTM-XGBoost model. *Solar Energy*, *300*, 113819. https://doi.org/10.1016/J.SOLENER.2025.113819