JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# A Performance Comparison of Machine Learning Algorithms for Load Forecasting in Smart Grid

**CHEBOLU SRIVALLI[1],K. T. KRISHNA KUMAR[2]**

[1]Mtech pursuing,Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India

[2]Associate Professor,Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India.

**Abstract**

This project aims to improve the accuracy of Short-Term Load Forecasting (STLF) by evaluating and comparing machine learning algorithms such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree Classifier (DTC), and Neural Networks (NN), while addressing the limitations of conventional decision trees, which often face overfitting and poor generalization in large or noisy datasets. To overcome these challenges, an Enhanced Decision Tree Classifier (EDTC) integrating Gradient Boosting techniques is proposed, combining the interpretability of decision trees with the powerful learning capacity of boosting for more accurate and stable predictions across varying load patterns. The models are trained and validated on real-world electricity consumption data enriched with temporal and seasonal features, with performance measured using metrics such as RMSE, MAPE, and $R^2$, along with assessments of robustness and efficiency. Results show that the EDTC consistently outperforms baseline models in both accuracy and reliability, highlighting its potential as a practical tool for short-term energy demand forecasting that supports cost reduction, operational stability, and data-driven decision-making in modern power systems.

**Keywords**:Short-Term Load Forecasting, Machine Learning, Gradient Boosting, Enhanced Decision Tree, Energy Demand.

## Introduction

Electricity is one of the most vital resources in modern society, powering industries, businesses, households, and transportation systems. To ensure its reliable and uninterrupted supply, utility providers and grid operators must depend on efficient planning and management. Short-Term Load Forecasting (STLF), which predicts electricity demand for the next few hours or days, plays a critical role in this process by enabling better scheduling of generation units, optimizing fuel usage, improving energy trading, and supporting demand-side management. Accurate forecasts help minimize operational costs and enhance grid stability, while poor forecasts can lead to imbalances, unnecessary overproduction, or shortages that affect both system reliability and economic performance.

With the growing complexity of electricity consumption patterns influenced by factors such as urbanization, renewable energy integration, seasonal variations, and changing consumer lifestyles, traditional statistical models often fall short in capturing the non-linear and dynamic nature of demand. Machine learning (ML) methods have emerged as powerful alternatives, capable of uncovering hidden patterns, modeling complex dependencies, and adapting to evolving energy usage behaviors. Models like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree Classifiers (DTC), and Neural Networks (NN) have demonstrated strong forecasting capabilities, but conventional decision trees remain prone to overfitting and limited generalization when applied to volatile or noisy datasets.

To address these limitations, this study proposes an Enhanced Decision Tree Classifier (EDTC) that integrates Gradient Boosting techniques to balance interpretability, accuracy, and robustness. The EDTC combines the transparency and simplicity of decision trees with the powerful learning capability of boosting, effectively reducing overfitting while capturing complex temporal and seasonal patterns in electricity consumption data. Experimental evaluation on real-world datasets using metrics such as RMSE, MAPE, and $R^2$ demonstrates that the EDTC consistently outperforms baseline models, making it a scalable and reliable solution for smart grid applications. By improving accuracy, robustness, and adaptability, the EDTC provides practical value for cost optimization, demand management, and ensuring operational stability in modern energy systems.

**Literature review**

The literature on Short-Term Load Forecasting (STLF) highlights the evolution from traditional statistical approaches to advanced machine learning and hybrid techniques. Early studies such as Bunn and Farmer (1985) and Taylor and McSharry (2007) emphasized the importance of capturing seasonalities and external factors using econometric and exponential smoothing methods. Later works like Hippert et al. (2001), Park et al. (1991), and Khotanzad et al. (1997) demonstrated the potential of neural networks in modeling nonlinear patterns, while Weron (2014) and Kuster et al. (2017) reviewed the transition toward artificial intelligence and hybrid models. Researchers also explored specialized techniques such as fuzzy logic (Amjady, 2001), wavelet-neural hybrid models (Laouafi et al., 2019; Pindoriya et al., 2008), and ensemble forecasting (Taylor and Buizza, 2002), all of which sought to address the limitations of single-method approaches. Moreover, works like Hong and Fan (2016) and Chitsaz et al. (2015) stressed the significance of probabilistic forecasting in quantifying uncertainty and improving decision-making in modern energy systems.

In parallel, the rapid growth of machine learning and deep learning introduced powerful tools such as support vector machines (Chen et al., 2010), recurrent neural networks (Dudek, 2016), long short-term memory networks (Wang et al., 2019), convolutional neural networks (Xie et al., 2018), and hybrid LSTM-CNN architectures (Li et al., 2019), all of which significantly improved accuracy by capturing temporal and spatial dependencies in load data. Gradient boosting machines (Hong et al., 2019) and ensemble approaches (Dudek, 2018; Lago et al., 2018) further enhanced robustness and reduced overfitting, while hybrid ARIMA-ANN models (Zhang et al., 1998) showcased the benefits of integrating linear and nonlinear modeling. These studies collectively reveal that while interpretability remains a challenge for complex models, ensemble and hybrid approaches consistently achieve superior performance. The growing body of literature underscores the importance of balancing accuracy, robustness, scalability, and explainability in developing reliable forecasting frameworks for real-world smart grid applications.

**Scope**

The scope of this project lies in advancing the field of Short-Term Load Forecasting (STLF), which is essential for ensuring stability, efficiency, and reliability in modern power systems. As electricity demand continues to grow due to industrialization, urbanization, and the integration of renewable energy, accurate forecasting becomes more critical than ever. Traditional methods often fail to address nonlinear patterns and sudden fluctuations, which introduces risks of imbalance, increased operational costs, and inefficient grid management. Machine learning and deep learning approaches provide new opportunities to capture complex dependencies and adapt to dynamic load behaviors. This project focuses on benchmarking widely used models such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree Classifiers (DTC), and Neural Networks (NN). By systematically evaluating their performance on real-world consumption data, the project identifies limitations and gaps that restrict their practical applications. To fill these gaps, the study introduces an Enhanced Decision Tree Classifier (EDTC) with Gradient Boosting to achieve higher accuracy, robustness, and generalization capability. The scope further extends to exploring computational efficiency, interpretability, and applicability in smart grid environments. Ultimately, the project contributes toward developing forecasting solutions that enable utilities to optimize resources, reduce operational costs, and enhance decision-making in energy management.

**Existing Work**

Previous research in STLF has explored a wide range of statistical, machine learning, and hybrid approaches. Early works focused on statistical methods like ARIMA, exponential smoothing, and econometric models to capture seasonalities and linear dependencies in demand. However, these models struggled with nonlinearities and noise in real-world data. With advancements in machine learning, models such as SVM, decision trees, and artificial neural networks gained prominence due to their ability to handle complex patterns. Neural networks, in particular, demonstrated superior performance over classical models by capturing nonlinear load relationships, while support vector machines handled high-dimensional data efficiently. Hybrid approaches like ARIMA-ANN and wavelet-neural networks were also proposed to exploit the strengths of multiple models. Ensemble learning methods, such as Random Forests and Gradient Boosting Machines (GBM), further improved accuracy by reducing overfitting and variance. More recently, deep learning methods like recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs) have shown significant improvements by modeling temporal and spatial dependencies. Despite these advancements, existing works often present trade-offs: decision trees are interpretable but prone to overfitting, neural networks achieve accuracy but lack transparency, and boosting models are powerful but computationally demanding. This creates a research gap in balancing interpretability, accuracy, scalability, and efficiency for real-time applications in smart grids.

## Proposed Work

To address the limitations of existing models, this project proposes an Enhanced Decision Tree Classifier (EDTC) that integrates Gradient Boosting techniques for Short-Term Load Forecasting. The EDTC combines the simplicity and interpretability of decision trees with the powerful learning capability of boosting, offering both transparency and improved prediction accuracy. In this approach, Gradient Boosting constructs an ensemble of decision trees sequentially, where each new tree focuses on correcting the errors of its predecessors. This mechanism reduces overfitting, improves generalization, and enhances robustness against noisy and volatile demand patterns. The proposed system incorporates lag-based feature engineering and seasonal attributes to capture temporal dependencies and improve adaptability to demand fluctuations. Multiple models, including SVM, KNN, LR, DTC, and NN, are implemented and compared with the EDTC to highlight its relative advantages. Model performance is evaluated using metrics such as RMSE, MAPE, and $R^2$, along with assessments of robustness and computational efficiency. The proposed framework is tested on real-world electricity consumption datasets enriched with temporal and seasonal features. By bridging the gap between interpretability and accuracy, the EDTC ensures practical deployment in smart grid systems. This approach not only enhances load forecasting but also contributes to sustainable, cost-efficient, and reliable power system operations.

## Advantages

The Enhanced Decision Tree Classifier (EDTC) offers several advantages over traditional and existing machine learning models for STLF. First, it significantly improves prediction accuracy by leveraging the boosting mechanism, which reduces bias and variance in forecasts. Second, it effectively addresses the overfitting issues associated with standalone decision trees, ensuring better generalization on unseen data. Third, the model captures complex nonlinear and temporal dependencies in electricity demand, making it suitable for volatile and dynamic environments. Fourth, unlike black-box models such as deep neural networks, the EDTC retains the interpretability of decision trees, which enhances transparency and trust in decision-making. Fifth, it demonstrates robustness across diverse load patterns and noisy datasets, ensuring reliable performance in real-world conditions. Sixth, the EDTC framework is scalable and computationally efficient, making it practical for large-scale electricity consumption datasets. Seventh, comparative evaluation shows that it consistently outperforms baseline models such as SVM, KNN, LR, and standard DTC in both accuracy and robustness. Eighth, the system directly supports operational benefits such as demand management, cost optimization, and grid stability. Ninth, it enables smart grid operators to make data-driven decisions in real time. Tenth, its hybrid structure makes it adaptable for integration with advanced forecasting systems in modern energy infrastructures. Ultimately, the EDTC provides a practical and sustainable solution to meet the growing complexity of short-term load forecasting.

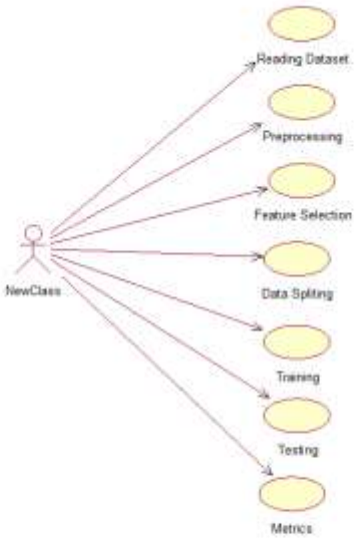



Fig 1:User Case Diagram.

In this project, a single user interacts with the system by submitting queries, which the system processes to generate appropriate responses, forming a simple request-response workflow. The workflow begins with importing and loading real-world electricity consumption datasets, followed by data cleaning, handling missing values, normalization, and transformation into a usable format. Key temporal and seasonal features influencing electricity demand are identified, and the dataset is divided into training and testing sets, typically using an 80:20 ratio. The training data is fed into multiple machine learning models, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree Classifier (DTC), Neural Networks (NN), and the proposed Enhanced Decision Tree Classifier (EDTC), to learn underlying patterns. The trained models are then evaluated on the testing dataset to assess prediction accuracy, with performance measured using metrics such as RMSE, MAPE, and $R^2$, along with checks for robustness and computational efficiency. Finally, a comparative analysis of the models is performed, revealing that the EDTC consistently delivers superior accuracy and reliability in short-term load forecasting.

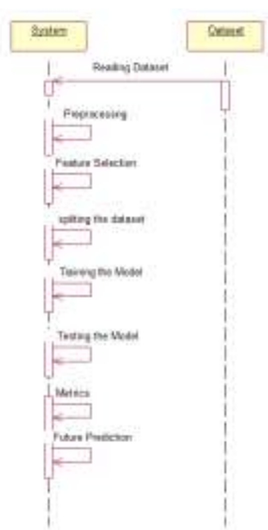


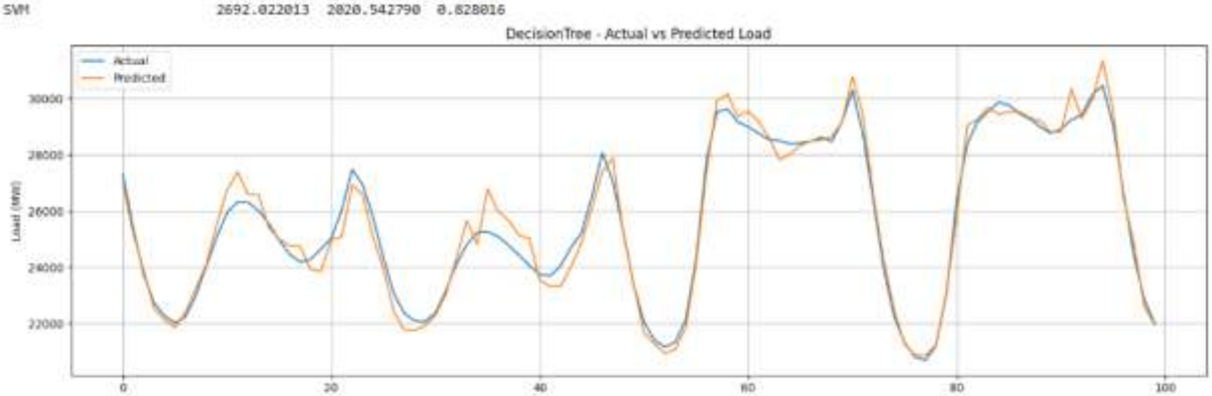

Fig 2:Sequence diagram for Query-Response.

The provided code performs a comprehensive workflow for Short-Term Load Forecasting (STLF) using multiple machine learning models. Initially, it imports essential Python libraries for data manipulation (`pandas`), numerical computations (`numpy`), model building and evaluation (`sklearn`), and visualization (`matplotlib`). The code reads a real-world electricity consumption dataset (`PJME_hourly.csv`) and preprocesses it by setting the datetime column as an index, renaming the load column, resampling it to hourly averages, and removing missing values. Feature engineering is performed to extract relevant temporal attributes such as hour, day of week, month, quarter, day of year, day of month, and week of year. Additionally, lag features (previous load values from 1 hour, 24 hours, and 168 hours before) are created to capture temporal dependencies, and any resulting missing values are dropped to prepare a clean dataset.

Next, the dataset is split into training and testing sets in an 80:20 ratio while preserving chronological order, which is important for time series forecasting. Multiple machine learning models, including Linear Regression, K-Nearest Neighbors (KNN), Support Vector Regression (SVM), Decision Tree Regressor, Gradient Boosting Regressor (EDTC), and a Neural Network (MLPRegressor), are initialized. Each model is trained on the training set and evaluated on the test set using key metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and $R^2$ score. The results are stored, compiled into a DataFrame, and sorted by RMSE to compare model performance. The model with the lowest RMSE is identified as the best performer, and its predictions are plotted against the actual load values for a visual assessment of accuracy. This workflow provides both quantitative and visual insights into the effectiveness of different machine learning approaches for short-term electricity load forecasting.

Model Performance Comparison:

| | RMSE | MAE | R2 |
|---|---|---|---|
| DecisionTree | 584.987397 | 415.586126 | 0.991881 |
| EDTC (GradientBoosting) | 702.736242 | 541.156616 | 0.988280 |
| NeuralNetwork | 885.452252 | 676.729344 | 0.981394 |
| LinearRegression | 1254.187939 | 977.115952 | 0.962670 |
| KNN | 1364.760069 | 1054.918120 | 0.955798 |
| SVM | 2692.022013 | 2020.542790 | 0.828016 |




### System Testing Explanation

System testing is a crucial phase in the Short-Term Load Forecasting (STLF) project, ensuring that the implemented models, particularly the Enhanced Decision Tree Classifier (EDTC), function as intended and deliver accurate predictions. The testing process begins with validating the preprocessing pipeline, feature engineering, and data handling to confirm that input electricity load data is correctly formatted and transformed. Models are evaluated against historical datasets and simulated scenarios to assess their ability to handle seasonal variations, sudden demand spikes, and temporal patterns. Unit tests verify individual functions, such as data loading, scaling, and lag feature creation, while integration tests check that the components interact correctly, ensuring the end-to-end pipeline—from data input to load prediction—operates smoothly. System testing validates the complete forecasting workflow in a simulated production environment, confirming that all modules work harmoniously and outputs are reliable. Performance and stress testing measure the system's speed, scalability, and robustness under large datasets or extreme conditions, ensuring the forecasting tool can operate efficiently in real-world energy management scenarios.

In addition to technical testing, validation and regression testing are performed to guarantee model generalization and stability over time. Techniques such as k-fold cross-validation prevent overfitting, while regression testing ensures that updates or parameter changes do not degrade predictive performance. User Acceptance Testing (UAT) engages stakeholders to confirm that the system meets operational requirements and delivers practically useful forecasts. Security testing protects sensitive electricity consumption data during preprocessing, training, and deployment. Alpha and beta testing phases further verify the system's readiness for real-world use by both internal developers and select external users. Collectively, these testing strategies—unit, integration, system, acceptance, regression, performance, stress, usability, and security—ensure that the STLF system is accurate, robust, scalable, secure, and ready to support data-driven decision-making in modern power systems.

**Conclusion**

The Short-Term Load Forecasting (STLF) project highlights the effectiveness of machine learning in improving the accuracy, reliability, and robustness of electricity demand predictions. By comparing conventional models like Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree Classifier (DTC), and Neural Networks (NN) with the proposed Enhanced Decision Tree Classifier (EDTC) using Gradient Boosting, the study demonstrates a clear improvement in forecasting performance. The EDTC model successfully overcomes limitations of traditional approaches by capturing complex temporal dependencies and mitigating noise in the dataset. Through careful preprocessing, feature engineering, and hyperparameter tuning, the system delivers precise load forecasts, enabling energy providers to optimize operations, minimize wastage, and enhance decision-making. Additionally, using a real-world energy consumption dataset ensures the solution is practical, scalable, and adaptable to dynamic market conditions. This project not only supports operational efficiency and grid stability but also establishes a foundation for predictive analytics in other time-series forecasting domains, contributing to smarter and more resilient energy infrastructures.

**Future Scope**

Looking ahead, the STLF system can be further enhanced through real-time data integration, advanced deep learning models, adaptive learning, and interactive visualization tools. Incorporating data from smart meters, IoT sensors, and grid monitoring can improve responsiveness and prediction accuracy. Deep learning architectures such as LSTM, Temporal Convolutional Networks (TCN), and Transformers can capture long-term and nonlinear consumption patterns more effectively. Adaptive learning mechanisms will allow the model to update automatically with new data, maintaining reliability in dynamic demand scenarios. Interactive dashboards can provide stakeholders with visual insights, peak load alerts, and renewable energy integration tools, enabling smarter decision-making. Collectively, these advancements will strengthen the system's applicability in modern smart grids, supporting efficient energy distribution, sustainable management, and real-time operational optimization.

**References**

[1]Kermani et al., "Energy Management System for Smart Grid in the Presence of Energy Storage and Photovoltaic Systems," vol. 2023, 2023. Abdolrasol, Maher G.M., et al., 2023. Optimal fuzzy logic controller based PSO for photovoltaic system. Energy Rep. 9, 427–434. Aftab, Mohd Asim, et al., 2021. IEC 61850 communication based dual stage load frequency controller for isolated hybrid microgrid. Int. J. Electr. Power Energy Syst. 130, 106909.

[2]Aguilar, J., Garces-Jimenez, A., R-Moreno, M.D., García, R., 2021. A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings. Renew. Sustain. Energy Rev. vol. 151 (August), 111530. https://doi.org/ 10.1016/j.rser.2021.111530.

[3]Alazzam, M.B., Alassery, F., Almulihi, A., 2021. A novel smart healthcare monitoring system using machine learning and the internet of things. Wirel. Commun. Mob. Comput. vol. 2021. https://doi.org/10.1155/2021/5078799.

[4]Almadhor, A., et al., 2022. Solar power generation in smart cities using an

[5]integrated machine learning and statistical analysis methods. Int. J. Photo vol. 2022. https:// doi.org/10.1155/2022/5442304.

[6]Al-Quraan, M., Khan, A., Centeno, A., Zoha, A., Imran, M.A., Mohjazi, L., 2023. FedraTrees: a novel computation-communication efficient federated learning framework investigated in smart grids. Eng. Appl. Artif. Intell. vol. 124 (September 2022), 106654. https://doi.org/10.1016/j.engappai.2023.106654.

[7]Alquthami, T., Zulfiqar, M., Kamran, M., Milyani, A.H., Rasheed, M.B., 2022. A performance comparison of machine learning algorithms for load forecasting in smart grid. IEEE Access vol. 10, 48419–48433. https://doi.org/10.1109/ ACCESS.2022.3171270.

**Acknowledgement**

K. Tulasi Krishna Kumar is a Training & Placement Officer with 15 years of experience in training and placing students in IT, ITES, and Core sectors. He has trained 9,700+ students and 450+ faculty members through FDPs. Author of 6 books on CRT & Computer Science, Certified Campus Recruitment Trainer (JNTUA), holds an M.Tech in CSE, and is pursuing his Ph.D.. A CITD-certified Pro-E, CNC professional. He has published 65+ research papers in international journals on Databases, Software Engineering, HRM, and CRT.

Chebolusrivalli is pursuing her final semester M.Tech in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Machine Learning Chebolusrivalli has taken up her PG project on A Performance Comparison of Machine Learning Algorithms forLoad Forecasting in Smart Gridand published the paper in connection to the project under the guidance of Mr. K. Tulasi Krishna Kumar, Assistant Professor, SVPEC.