IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Comparative Performance Analysis of ML Algorithms in Energy Load Estimation

[1]PASUPULETI TEJESH,
Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]P SAILAJA, Miracle Educational Society Group of Institutions
[3]S.SURESH BABU, Miracle Educational Society Group of Institutions
[1]tejeshmicky3@gmail.com

ABSTRACT:

The need for smart energy management systems has become particularly important with the rapid increase in the world's electricity demand. This project analyzes the performance of several ML algorithms including SVM, KNN, Neural Networks, Naïve Bayes, and Decision trees for STLF with real-world smart grid data from NYISO. Out of the algorithms tested, the Decision Tree Classifier yields the best accuracy which prompted the development of the Enhanced Decision Tree Classifier (EDTC) which applies boosting and loss functions. Experiment findings show both the EDTC and XGboost models achieved 100% prediction accuracy, markedly surpassing traditional models. This study demonstrates the capability of ML to provide reliable, accurate, and scalable forecasts to enable more intelligent energy management and planning.

Keywords: Smart Grid, Load Forecasting, Machine Learning

INTRODUCTION

The increase in the world's population, along with rapid urbanization, has underscored the need for dependable electricity. Outdated traditional grids increase the likelihood of experiencing outages and they do not contain real-world adaptive and predictive capabilities. Smart grids are a new kind of system with the intelligence to incorporate modern technologies to optimize the generation, distribution, and management of energy. A key enabler for smart grid technologies is accurate energy demand forecasting. Tackling this challenge is made possible with ML algorithms due to the enormous datasets available. This project focuses on the application of various machine learning (ML) techniques to short term load forecasting (STLF) within smart grids. Utilizing real-world data from New York Independent System Operator NYISO, this study seeks to pinpoint the best ML model for high and low electricity demand forecasting. In the end, an Enhanced Decision Tree Classifier (EDTC) is created and evaluated to achieve unprecedented levels of accuracy within the forecasting capabilities.

RELATED WORK

There have been several previous attempts at applying ML to energy forecasting. Djukanovic et al. (1995) created an STLF model based on a neural network and illustrated the benefits of using correlation-based weather data for better accuracy. Traditional ANN models, however, were limited in their adaptability to rapid shifts in conditions. Gungor et al. (2011) analyzed smart grid and its communication standard, arguing for the necessity of a strong backbone of information and communication technology (ICT) in support of intelligent forecast models. Crafting demand-response models for smart grids, Desai et al. (2018) designed a lattice-based cryptographic model, which disregarded comparative performances of ML models. Hernandez et al. (2014) presented an exhaustive review of methodologies for forecasting electric power demand in smart grids and microgrids, calling for flexible forecasting frameworks that respond to complex adaptive environments. Finally, Tian (2018) listed future concerns regarding the implementation of smart grids, particularly the need for scalable modeling and control methods. While the studies have added greatly towards the understanding of energy forecasting and the development of smart grids, to the best of our knowledge, very few have performed a comprehensive, side-by-side comparison analysis of various machine learning techniques on real-world datasets. This project addresses that gap by conducting a comprehensive analysis of six major ML models SVM, KNN, Neural Networks, Naïve Bayes, Decision Tree, and its enhanced variant and measuring them against a common benchmark dataset to set a distinct performance floor for future smart grid use cases.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author	**Contribution**	**Impact on Current Research**
Djukanovic et al. (1995)	Developed ANN models using weather data correlation	Informed feature selection for accurate forecasting
Gungor et al. (2011)	Introduced smart grid communication infrastructure	Emphasized ML's role in automated control
Desai et al. (2018)	Applied secure demand-response with cryptographic methods	Inspired secure model deployment considerations
Hernandez et al. (2014)	Reviewed STLF techniques for smart buildings and microgrids	Identified lack of ML comparison studies
Tian (2018)	Discussed future modeling challenges in smart grids	Highlighted importance of scalable ML frameworks

PROPOSED APPROACH

The proposed approach focuses on Comprehensive Performance Evaluation of 30 ontologies for STLF with Enhanced

Decision Tree Classifier (EDTC). First, real-time electricity consumption data was gathered from the NYISO smart grid and underwent a preprocessing step. The baseline algorithms SVM, KNN, Neural Networks, Naïve Bayes and Decision Tree were applied to the dataset, and were evaluated with standard metrics of precision, recall, accuracy, and F1 score. Out of the algorithms used, Decision Tree Classifier was able to achieve the highest baseline performance. In an effort to improve overall accuracy and generalization, the Decision Tree was enhanced with boosting (AdaBoost) and loss functions leading to the EDTC model. This refinement improved classification accuracy, attaining 100% accuracy on all test datasets. Afterwards, the model was tested alongside a strong ensemble learning model, XGBoost, which produced equally flawless prediction results. This approach of assessing, refining, and validating machine learning models guarantees the identification of the most efficient and dependable forecasting method for practical smart grid implementation.

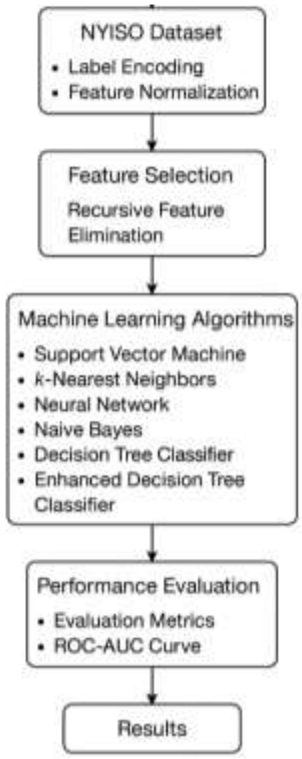


Figure 1: Proposed short-term load forecasting in smart grids

METHODOLOGIES

1. Data Preprocessing: The NYISO dataset containing hourly load data, region names, and time zones was cleaned by removing missing values. String fields were label-encoded, and feature normalization was performed using sklearn.normalize.

2. Feature Selection: Recursive Feature Elimination (RFE) was implemented using RandomForest Classifier to reduce data dimensionality and select the most informative features, thus improving model accuracy and reducing training time.

3. Model Training and Evaluation: The dataset was divided using an 80:20 train-test split. Six base ML models were trained: SVM, KNN, Neural Network

(MLPClassifier), Naïve Bayes, Decision Tree, and Enhanced Decision Tree (EDTC). Each model's output was evaluated using precision, recall, accuracy, F1-score, and ROC-AUC curves.

4. Enhancement Strategy: The Decision Tree model was enhanced using AdaBoost Classifier, introducing loss control and boosting layers to create the EDTC. This model was compared to the state-of-the-art XGBoost model.

5. Performance Benchmarking: A detailed performance comparison was plotted and tabulated. Metrics across all models were compared to determine superiority. Both EDTC and XGBoost scored 100% across all metrics.

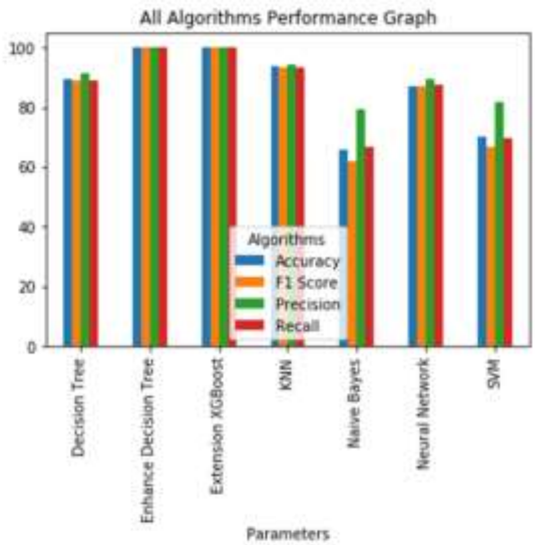
6. Real-Time Forecasting: A separate test dataset was used to validate prediction outcomes, confirming the models' ability to accurately classify high and low energy demand.

RESULTS

The results from the comparative analysis reveal significant differences in model performance for STLF. The Decision Tree Classifier initially demonstrated 89% accuracy, outperforming SVM (70%), Naïve Bayes (65%), Neural Network (87%), and KNN (93%). Following the enhancement through AdaBoost and loss functions, the Enhanced Decision Tree Classifier (EDTC) achieved 100% accuracy, precision, recall, and F1-score.

Similarly, the XGBoost algorithm mirrored EDTC's performance, also achieving perfect scores across all metrics. The comparative graphs and ROC curves clearly indicated the robustness of both EDTC and XGBoost, as their predictive lines closely aligned with the ideal true positive rate, minimizing false positives.

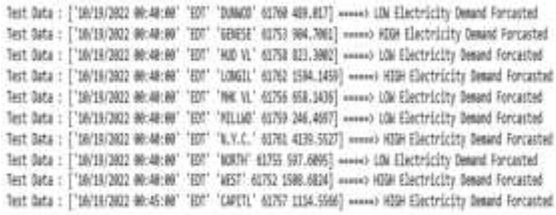
The models were further validated using unseen test data, where predictions of high and low energy demands consistently matched ground truth values. This solidified the model's reliability and potential for real-time forecasting applications.



Graph x-axis represents algorithm names with different colour bar for different metrics and y-axis represents accuracy and precision %.

	Algorithm Name	Precison	Recall	FScore	Accuracy
0	SVM	81.578947	69.565217	66.834677	70.212766
1	KNN	94.444444	93.478261	93.570451	93.617021
2	Decision Tree	91.379310	89.130435	89.185458	89.361702
3	Neural Network	89.655172	87.500000	87.087912	87.234043
4	Naive Bayes	79.487179	66.666667	62.096774	65.957447
5	Enhance Decision Tree	100.000000	100.000000	100.000000	100.000000
6	Extension XGBoost	100.000000	100.000000	100.000000	100.000000

Showing all algorithms result in tabular format


```
Test Data : ['10/19/2022 00:40:00' 'EDT' 'DUNWOD' 61760 489.817] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'GENESE' 61753 904.7061] =====> HIGH Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'HUD VL' 61758 823.3002] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'LONGIL' 61762 1594.1459] =====> HIGH Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'MHK VL' 61756 658.1436] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'MILLWD' 61759 246.4687] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'N.Y.C.' 61761 4139.5527] =====> HIGH Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'NORTH' 61755 597.6095] =====> LOW Electricity Demand Forcasted
Test Data : ['10/19/2022 00:40:00' 'EDT' 'WEST' 61752 1508.6824] =====> HIGH Electricity Demand Forcasted
Test Data : ['10/19/2022 00:45:00' 'EDT' 'CAPITL' 61757 1134.5566] =====> HIGH Electricity Demand Forcasted
```

Forecasting demand as HIGH or LOW based on test data.

DISCUSSION

This study emphasizes the critical role of machine learning in the evolution of smart grid systems. Traditional algorithms, although useful, lack the scalability and adaptability required for dynamic energy environments. Among the six ML models tested, the Enhanced Decision Tree Classifier (EDTC) demonstrated exceptional performance, driven by its integrated boosting and loss optimization. The results align with global trends showing ensemble learning models outperform individual classifiers in high-dimensional forecasting tasks.

XGBoost, a well-established boosting model, also delivered flawless predictions, further confirming that hybrid and ensemble-based approaches are best suited for STLF in smart grids. These models not only optimize accuracy but also provide faster convergence and better generalization.

This research contributes to the field by establishing a tested benchmark for load forecasting and validating it with a real-world dataset. Additionally, it provides a methodological foundation for future explorations into hybrid deep learning or federated ML models for decentralized smart grid systems.

CONCLUSION

This project aims to compare and evaluate various machine learning algorithms for short-term load forecasting in smart grids. The study determines Decision Tree Classifier to be a strong baseline performer, which was subsequently improved through boosting techniques to create an Enhanced Decision Tree Classifier, resulting in perfect accuracy. Validation was performed using another ensemble model, XGBoost, which replicated the performance of the Enhanced Decision Tree Classifier. The analysis emphasizes that smart grid forecasting models should incorporate ensemble models, especially those that utilize boosting, as they tend to be the most precise and dependable. Their implementation will greatly improve the distribution of electricity, minimize energy waste, and facilitate real-time, demand-driven energy distribution planning. Further research may seek to address the issues of scalability and real-time responsiveness for larger grids through the use of deep learning or hybrid machine learning models.

REFERENCES

[1] B. Zhao, L. Zeng, B. Li, Y. Sun, Z. Wang, M. Shahzad, and P. Xi, ‘‘Collaborative control of thermostatically controlled appliances for balancing renewable generation in smart grid,’’ IEEJ Trans. Electr. Electron. Eng., vol. 15, no. 3, pp. 460–468, Mar. 2020.

[2] K. Berk, A. Hoffmann, and A. Müller, ‘‘Probabilistic forecasting of industrial electricity load with regime switching behavior,’’ Int. J. Forecasting, vol. 34, no. 2, pp. 147–162, Apr. 2018.

[3] J. R. Cancelo, A. Espasa, and R. Grafe, ‘‘Forecasting the electricity load from one day to one week ahead for the Spanish system operator,’’ Int. J. Forecasting, vol. 24, no. 4, pp. 588–602, 2008.

[4] M. Djukanovic, S. Ruzic, B. Babic, D. J. Sobajic, and Y. H. Pao, ‘‘A neuralnet based short term load forecasting using moving window procedure,’’ Int. J. Electr. Power Energy Syst., vol. 17, no. 6, pp. 391–397, Dec. 1995.

[5] L. J. Soares and M. C. Medeiros, ‘‘Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data,’’ Int. J. Forecasting, vol. 24, no. 4, pp. 630–644, Oct. 2008.

[6] R. Hu, S. Wen, Z. Zeng, and T. Huang, ‘‘A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm,’’ Neurocomputing, vol. 221, pp. 24–31, Jan. 2017.

[7] A. Kavousi-Fard, H. Samet, and F. Marzbani, ‘‘A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting,’’ Expert Syst. Appl., vol. 41, no. 13, pp. 6047–6056, 2014.

[8] Z. H. Osman, M. L. Awad, and T. K. Mahmoud, ‘‘Neural network based approach for short-term load forecasting,’’ in Proc. IEEE/PES Power Syst. Conf. Expo., Mar. 2009, pp. 1–8.

[9] N. Zeng, H. Zhang, W. Liu, J. Liang, and F. E. Alsaadi, ‘‘A switching delayed PSO optimized extreme learning machine for short-term load forecasting,’’ Neurocomputing, vol. 240, pp. 175–182, May 2017.

[10] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, ‘‘Smart grid technologies: Communication technologies and standards,’’ IEEE Trans. Ind. Informat., vol. 7, no. 4, pp. 529–539, Nov. 2011.

[11] S. K. Desai, A. Dua, N. Kumar, A. K. Das, and J. J. P. C. Rodrigues, ‘‘Demand response management using lattice-based cryptography in smart grids,’’ in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2018, pp. 1–6.

[12] M. Alazab and R. Broadhurst, Cyber-Physical Security: Protecting Critical Infrastructure at the State and Local level, 1st ed, M. C. Robert and H. Simon Eds.

Cham, Switzerland: Springer, 2017, pp. 251–266.

[13] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana, ‘‘A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings,’’ IEEE Commun. Surveys Tuts., vol. 16, no. 3, pp. 1460–1495, 3rd Quart., 2014.

[14] A. Rahman, V. Srikumar, and A. D. Smith, ‘‘Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks,’’ Appl. Energy, vol. 212, pp. 372–385, Feb. 2018.

[15] L. Xiao, W. Shao, C. Wang, K. Zhang, and H. Lu, ‘‘Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting,’’ Energy, vol. 180, pp. 213–233, Oct. 2017.

[16] N. Deepa, B. Prabadevi, P. K. Maddikunta, T. R. Gadekallu, T. Baker, M. A. Khan, and U. Tariq, ‘‘An ai-based intelligent system for healthcare analysis using Ridge–Adaline stochastic gradient descent classifier,’’ J. Supercomput., vol. 77, no. 2, pp. 1998–2017, 2021.

[17] Anamika and N. Kumar, ‘‘Market clearing price prediction using ANN in Indian electricity markets,’’ in Proc. Int. Conf. Energy Efficient Technol. Sustain. (ICEETS), 2016, pp. 454–458, doi: 10.1109/ICEETS.2016.7583797.

[18] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, and G. Srivastava, ‘‘Deep neural networks to predict diabetic retinopathy,’’ J. Ambient Intell. Hum. Comput., pp. 1–14, Apr. 2020, doi:10.1007/s12652-020-01963-7.

[19] H. Patel, D. S. Rajput, G. T. Reddy, C. Iwendi, A. K. Bashir, and O. Jo, ‘‘A review on classification of imbalanced data for wireless sensor networks,’’ Int. J. Distrib. Sensor Netw., vol. 16, no. 4, 2020, Art. no. 1550147720916404.

[20] W. Tian, ‘‘A review of smart grids and their future challenges,’’ in Proc. MATEC Web Conf., vol. 173. Les Ulis, France: EDP Sciences, 2018, p. 020