Explainable AI for Reliable Energy Consumption Forecasting: A SHAP-based Approach

Qianqian He
Hangzhou Dianzi University
Hangzhou, China
bellaland11@naver.com

***Abstract*—**Carbon reduction and energy conservation are pressing global issues, with many countries implementing policies like China's 2024-2025 Action Plan, the European Union's Fit for 55 package, and the United States' Inflation Reduction Act (IRA). Machine learning algorithms provide a valuable tool predicting energy consumption, which can support these energy conservation efforts. In this study, we evaluated multiple machine learning models on an energy consumption dataset. Our results show that XGBoost achieved the best predictive performance, as indicated by the lowest Mean Absolute Error (MAE) and Mean Squared Error (MSE), while Support Vector Machine (SVM) performed the worst. To further understand the differences in these models' predictions, we applied SHAP (Shapley Additive exPlanations). SHAP analysis revealed that the same features influenced XGBoost and SVM differently, contributing to the varying prediction outcomes. This approach not only highlights the importance of feature-specific behavior in different algorithms but also enhances the transparency and trustworthiness of machine learning models. Such explainable AI techniques can be applied in industries like manufacturing to improve the usability and reliability to power consumption prediction models.

Keywords-*Energy consumption prediction, Machine learning, XGBoost, SHAP (Shapley Additive exPlanations), Explainable AI*

I. INTRODUCTION

Energy has always been a pivotal topic in human development, driving both technological advancements and societal transformations. From the 1970s, when the global community first began recognizing the challenges of energy scarcity and environmental impact, to the present-day pursuit of net-zero emissions in the 2020s, there has been a persistent effort to increase energy efficiency and reduce power consumption [1]. Achieving these goals is vital for mitigating climate change and promoting sustainable development. One of the key technologies that can contribute to this effort is power consumption prediction, which enables more efficient energy use, facilitates energy-saving policies, and helps to manage energy resources more effectively.

Machine learning (ML) algorithms have emerged as powerful tools for predicting energy consumption due to their ability to analyze complex, nonlinear patterns in data. However, while these models can deliver highly accurate predictions, they often operate as "black boxes", making it difficult to understand how individual features contribute to the overall prediction. This lack of transparency can hinder trust in the model's reliability, particularly in applications where explainability is critical, such as energy management. To ensure machine learning algorithms are both trustworthy and reliable for power consumption forecasting, it is important not only to optimize their accuracy but also to understand how they derive their results [2].

In this study, we aim to address this gap by comparing various machine learning models to uncover the underlying factors that influence their prediction performance. Using the same energy consumption dataset, we apply identical data preprocessing steps – including data cleansing and feature engineering – to ensure that the only variable is the machine learning algorithm itself. By applying SHAP (SHapley Additive exPlanations), we provide an in-depth analysis of how different features contribute to the prediction in each algorithm. This comparative analysis reveals that the importance of features varies significantly between models, offering new insights into why prediction results differ.

II. LITERATURE REVIEW

Energy consumption prediction has attracted extensive attention in recent years due to the global focus on energy conservation and carbon reduction. Numerous machine learning models have been applied in this area, showing significant potential for improving prediction accuracy. For instance, Olu-Ajayi, et al. explored various machine learning algorithms – including Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees (DT) – to forecast energy consumption in residential buildings [3]. Their study demonstrated that deep learning models, such as Deep Neural Networks (DNN), were particularly effective in this context. Similarly, Zhao et al. proposed an extreme deep learning model combining stacked autoencoders with extreme learning machines to enhance the accuracy of building energy consumption predictions [4].

However, while machine learning algorithms have shown great promise in energy prediction, they are often criticized for their "black box" nature, which limits interpretability. This opacity raises concerns about trust and reliability, especially when these models are applied in critical domains such as energy management. To address these concerns, researchers have developed a range of explainable AI (XAI) methods designed to make machine learning models more transparent. Among the most prominent methods are SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). These techniques help clarify how individual features contribute to model predictions, thereby enhancing trust in machine learning algorithms [5, 6].

Several studies have specifically highlighted the importance of explainable AI in energy forecasting. For example, Lundberg and Lee’s development of SHAP provides a unified framework for interpreting model predictions, offering detailed insights into how different features influence model outputs [5]. By comparing SHAP with other XAI methods, such as LIME, research has shown that SHAP provides more consistent and holistic explanations, making it particularly suitable for complex models like those used in energy prediction. As a result, SHAP has been increasingly adopted in energy prediction studies to explain the feature importance in models like XGBoost and SVM [7].

In conclusion, while earlier work in energy prediction primarily focused on improving model accuracy, there is now a growing emphasis on the interpretability of machine learning models. SHAP has emerged as one of the most effective tools for achieving this goal, as it provides clear and consistent explanations of how different features contribute to predictions. This study builds on previous work by applying SHAP to compare the feature importance across various machine learning algorithms used for energy consumption prediction, thereby improving both the transparency and trustworthiness of these models.

III. RESEARCH METHODS

As shown in Figure 2, to predict power consumption and understand how it works, we start from data preparation. And then we use five machine learning methods to predict power consumption to get the best prediction result. Finally, to uncover why algorithm works, we use SHAPLEY to explore the underlying principles of each modelling method.

Figure 2. Flow Diagram of Our Research.

A. Data Preparation

This research begins with the preparation of power consumption data sourced from EXIST, Turkey’s official energy exchange authority, which provides historical electricity consumption data. The dataset records hourly electricity consumption (in MWh) for each day from January 1, 2016, to March 24, 2020. The data contains detailed time-series information that spans over four years, which is crucial for understanding long-term trends and seasonality in energy usage. We use Numpy and Pandas for extracting, transforming and loading data. And then visualizing data with Seaborn and Matplotlib to track power consumption over time as shown in Figure 1.

Feature selection maintains a subset of the original features, whereas feature extraction creates new features from the original dataset [13]. The machine learning and deep learning approaches developed in the last decade provide a very high level of accuracy of various types of applications, including time-series forecasting [14]. For time-series forecasting, 22 features in total are generated via feature engineering. Firstly, we capture the consumption from 38, 41, 48, 72, and 168 time steps ago to extract 5 lag features. Next, we shift the data by 38 time steps, then computing the standard deviation over the next 12, 24, and 48 time steps to extract 9 rolling features. Finally, we select hour, quarter, month, year, dayofyear, day, weekofyear, and days_in_month to generate 8 time features.

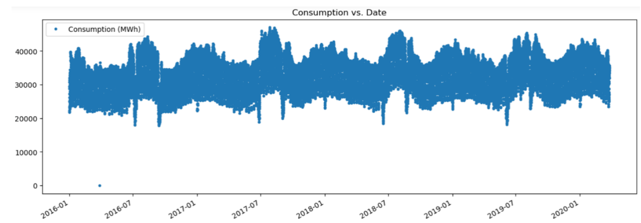


Figure 1. Time Series Time Series Plot (Consumption vs. Date).

Afterwards the data set is split into training set (data from 2016-01-01 to 2019-12-31, 34895 records) and test set (data from 2020-03-14 to 2020-03-24, 1776 records) as shown in Figure 3.

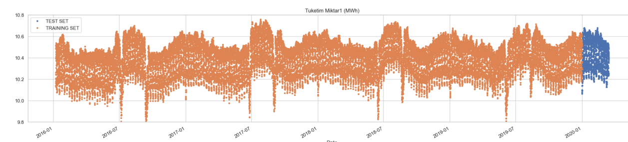


Figure 3. Time Series Plot of Test Set & Training Set.

B. Modeling

By now, the data is ready for modeling. In total, 5 kinds of machine learning algorithms as shown in Table I including random forest, decision tree, XGBoost, lightGBM, SVM are used for training and predicting power consumption.

Hyperparameter optimization is pivotal in machine learning for enhancing the perfor- mance of models. While models come with default hyperparameters, fine-tuning them to a specific dataset can substantially boost their efficacy [15].

TABLE I. DESCRIPTION OF MACHINE LEARNING ALGORITHMS

Name	Definition	Reference
Random Forest	an ensemble learning method that builds multiple decision trees during training	“Random forests” by Breiman [8]
Decision Tree	a non-parametric supervised learning method used for classification and regression	“Classification and Regression Trees” by Breiman, et al. [9]
XGBoost	an optimized gradient-boosting algorithm using an ensemble of weak learners	“XGBoost: A Scalable Tree Boosting System” by Chen and Guestrin [10]
LightGBM	a gradient-boosting framework that uses tree-based learning algorithms	“LightGBM: A Highly Efficient Gradient Boosting Decision Tree” by Ke, et al. [11]
SVM	a supervised learning algorithm that finds the hyperplane that best separates data points	“A Training Algorithm for Optimal Margin Classifiers” by Boser, et al. [12]

To compare the performance of different algorithms with the same dataset, we need to control variables. Therefore, in our research all algorithms are applied with original hyperparameter without tuning. By default, some algorithms like XGBoost have optimized hyperparameter for general purpose.

C. Evaluation

To evaluate the forecast error of time series prediction, we use MAE (Mean Absolute Error, the arithmetic average of the absolute errors) and MSE (Mean Squared Error, the average of the squares of the errors).

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} = \frac{\sum_{i=1}^{n}|e_i|}{n} \quad (1)$$

Where: n = number of data points. y_i = observed values.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (2)$$

Where: n = number of data points. Y_i = observed values. $\hat{Y}_i$ = predicted values

D. Explanation

As indicated by MSE and MCE of various prediction results, performance varies while using different algorithms. To explain why this happens, we use SHAP values and visualizations for every algorithm to see how each feature contributes to a specific prediction.

SHAP (SHapley Additive exPlanations) is a unified framework for interpreting predictions. Below formula computes the average marginal contribution of a feature i across all possible feature combinations. For model $f(x)$ and a feature subset S , the Shapley value for feature i is:

$$\emptyset_i(f) = \sum_{s \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(S \cup \{i\}) - f(S)] \quad (3)$$

Where: F: the set of all features. $f(S)$: the model prediction when only features in subset S are included.

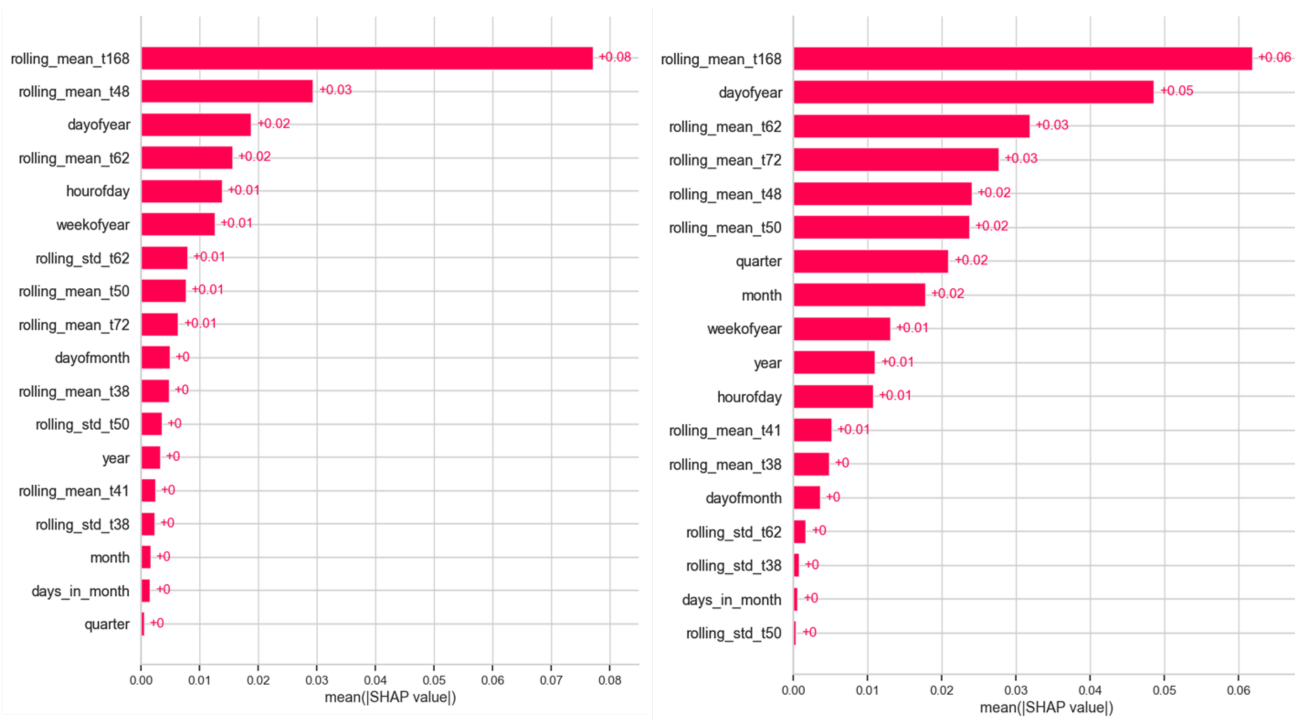
IV. EXPERIMENT

The experiment is programmed with Python and Tensorflow using Jupyter on the Mac Pro with M1 chip.

As shown in Table 2, XGBoost has the best accuracy, while SVM is the poorest.

TABLE II. MAE & MSE OF PREDICTION RESULTS

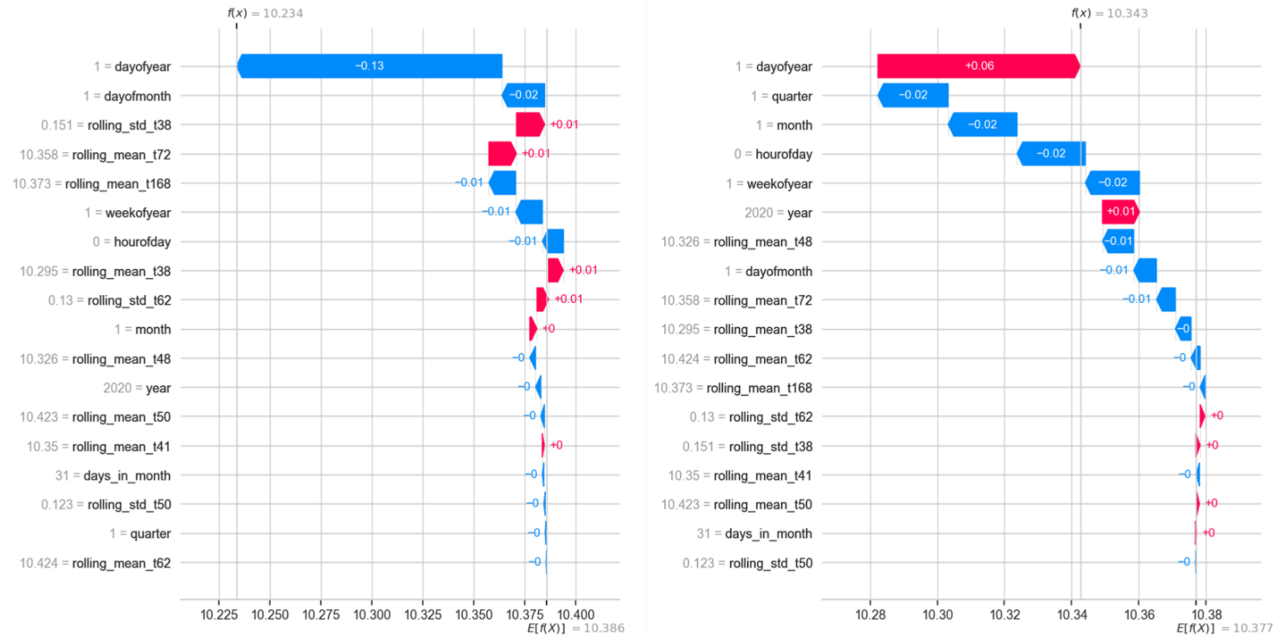
Algorithm	MAE	MSE
XGBoost	0.0296884	0.00169442
random forest	0.03237786	0.0025887
lightgbm	0.03256667	0.0022812
decision tree	0.04455723	0.00469673
svm	0.04520031	0.00428869



(a) Bar Plot of XG Boost

(b) Bar Plot of SVM

Figure 4. Bar Plot of XGBoost and SVM.



(a) Waterfall Plot of XG Boost (b) Waterfall Plot of SVM

Figure 5. Bar Plot of XGBoost and SVM.

To find out why XG Boost and SVM have different results predicting the same data set with the same features, we use SHAPLEY values and visualizations as below to understand the contribution of each feature to the model's prediction.

Bar Plot shows the importance of features by the mean absolute SHAP value for each feature. The influence of the feature is increased as the bar becomes higher. As shown in Figure 4, the most important features of XG Boost and SVM are different. For XG Boost, rolling_mean_t168, rolling_mean_t48, and dayofyear are the top3 features that contribute the most. While for SVM, the most important 3 features are rolling_mean_t168, dayofyear and rolling_mean_t62.

Waterfall Plot provides a clear explanation of which features are the most important for a given prediction. As shown in Figure 5, the features of pushing the prediction higher or lower than the baseline are different. For XG Boost, dayofyear pushes the prediction lower. While for SVM, dayofyear pushes the prediction higher.

V. CONCLUSION

In this study, we compared the predictive performance of several machine learning algorithms for power consumption forecasting. As indicated by the Mean Absolute Error (MAE) and Mean Squared Error (MSE), XGBoost demonstrated the highest accuracy, while Support Vector Machine (SVM) had the lowest predictive performance. This discrepancy can be further explained through SHAP (SHapley Additive exPlanations), which reveals that the same feature contributes differently to each model's predictions. The variance in feature importance across models highlights that each algorithm processes the data in unique ways, leading to variations in prediction outcomes.

Moreover, by using SHAP, we gain a deeper understanding of how individual features influence the model's predictions, thereby enhancing our trust in the algorithms. This explainability is crucial not only for evaluating performance but also for ensuring transparency in machine learning applications.

In practical terms, the findings suggest that explainable AI techniques like SHAP can be applied beyond energy consumption forecasting. For instance, in traditional industries such as manufacturing, explainable AI can improve both the reliability and usability of machine learning algorithms, making them more trusted tools for decision-making in real-world applications. By providing clear insights into how predictions are generated, explainable AI enhances the integration of machine learning into critical industry sectors, fostering more efficient and reliable predictions in various contexts.

REFERENCES

[1] IEA. "How Energy Efficiency Will Power Net Zero Climate Goals." https://www.iea.org/commentaries/how-energy-efficiency-will-power-net-zero-climate-goals (accessed)

[2] Q. Yin, C. Han, A. Li, X. Liu, and Y. Liu, "A Review of Research on Building Energy Consumption Prediction Models Based on Artificial Neural Networks," *Sustainability*, vol. 16, no. 17, p. 7805, 2024.

[3] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques," *Journal of Building Engineering*, vol. 45, p. 103406, 2022.

[4] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, "Building energy consumption prediction: An extreme deep learning approach," *Energies*, vol. 10, no. 10, p. 1525, 2017.

[5] S. Lundberg, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd*

ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135-1144.

[7] M. Christoph, *Interpretable machine learning: A guide for making black box models explainable*. Leanpub, 2020.

[8] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5-32, 2001.

[9] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, “Classification and Regression Trees,” *Biometrics*, vol. 40, p. 874, 1984.

[10] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.

[11] G. Ke *et al.*, “LightBGM: A highly efficient gradient boosting decision tree,” *Advances in neural information processing systems*, vol. 30, 2017.

[12] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144-152.

[13] Htet Htet Htun, Michael Biehl, and Nicolai Petkov, “Survey of feature selection and extraction techniques for stock market prediction,” in *International Journal of Computational Intelligence Systems*, vol. 9, no. 26, p. 7, 2023.

[14] Nebojsa Bacanin, Catalin Stoean, Miodrag Zivkovic, Miomir Rakic, Roma Strulak-Wójcikiewicz and Ruxandra Stoean, “On the Benefits of Using Metaheuristics in the Hyperparameter Tuning of Deep Learning Models for Energy Load Forecasting,” *Energies*, vol. 16, p. 1434, 2023.

[15] Medhi Imani, and Hamid Reza Arabnia, “Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction: A Comparative Analysis,” *Technologies*, vol. 11, p. 167, 2023.

2024 4th International Conference on Electronic Information Engineering and Computer Technology

11/15/2024 - 11/17/2024 Shenzhen•China

Acceptance Letter

Dear Author(s):

Congratulations! Your manuscript has passed the peer review (the reviewers' comments are available in the attached file on AIS) and has been accepted by the 2024 4th International Conference on Electronic Information Engineering and Computer Technology. The conference will be held in Shenzhen•China from 11/15/2024 - 11/17/2024. We are glad to invite you to attend the conference and make an oral report.

Manuscript No.: VZZRJ3ENEA

Author name(s): Qianqian He

Manuscript title: Explainable AI for Reliable Energy Consumption Forecasting: A SHAP-based Approach

Your manuscript, after presented in the oral report or poster in the conference, will be published on Conference Proceedings, after which it will be submitted for index in EI Compendex、Scopus.

2024 4th International Conference on Electronic Information Engineering and Computer Technology

The Committee of EIECT 2024

11-15-2024

Notices:

1. Authors need to revise the manuscript as per the reviewers' comments before re-uploading the final version (in Word or PDF) to the AIS system.
2. Authors need to ensure that the submitted manuscript is an original paper with a similarity lower than 20%. Once the manuscript is submitted to AIS, the authors are not allowed to re-submit it to other journals for publication.
3. Authors need to confirm their attendance one week before the conference is held. If the authors are not able to be present on the conference after agreeing to attend the conference, the authors need to reach the conference secretary for re-arrangement.

Again, congratulations and we look forward to meeting you in Shenzhen•China