Vokasi Unesa Bulletin of Engineering, Technology and Applied Science (VUBETA)
https://journal.unesa.ac.id/index.php/vubeta
Vol. 2, No. 3, 2025, pp. 549~557 DOI: 10.26740/vubeta.v2i3a.40135
ISSN: 3064-0768

Estimating Residential Natural Gas Demand and Consumption: A Hybrid Ensemble Machine Learning Approach

Mohammed Ajuji [1*], Ahmed Mohammed [2], Muhammad Dawaki[3], Abuzairu Ahmad[4]
[1,2,3]Department of Computer Science, Faculty of Science, Gombe State University (GSU), Gombe, Nigeria
[4]Department of Mathematical Sciences, Abubakar Tafawa Balewa University (ATBU), Bauchi, 740272, Nigeria

Article Info

Article history:

Received May 3, 2025
Revised June 18, 2025
Accepted July 22, 2025

Keywords:

Estimation
Natural gas
Hybrid
Ensemble
Machine learning

ABSTRACT

Natural gas is often used for cooking, drying clothes, heating, etc., particularly in residential settings; it has been an essential component for human beings for many decades. This study proposes a hybrid ensemble regression machine learning model for forecasting residential natural gas demand. Accurate demand prediction tends to reduce energy waste and address some of the energy challenges; such as the need for reliable, affordable, and sustainable energy consumption, thereby, improving energy management and resource planning. The proposed approach integrates multiple regression algorithms including K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Decision Tree Regression (DTR), and Linear Regression (LR) to leverage the strengths of each model to develop a hybrid model that enhances overall predictive performance. The ensemble method operates in two phases: training individual regression models on the dataset, followed by aggregating their predictions. Model performance is evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), coefficient of determination (R^2), and prediction accuracy, and is benchmarked against individual models. Cross-validation techniques were applied to ensure the robustness of the results. Experimental results demonstrate that the hybrid ensemble approach consistently outperforms standalone models by capturing diverse patterns and relationships within the data.

1. INTRODUCTION

The use of natural gas in various aspects of human endeavors, particularly in residential buildings, has been prevalent for several decades [1]. The only byproducts of burning natural gas are carbon dioxide, water vapor, and trace amounts of nitrogen oxide, making it the cleanest among fossil fuels on earth. A large number of consumer goods, such as stoves, dryers, fireplaces, and furnaces, are also powered by natural gas [2]. Home appliances, probably at least one of them, rely on natural gas. Natural gas, like most other energy sources, can be hazardous if not handled properly. It is wise and ideal to learn how to protect oneself and others by adhering to a few basic safety precautions and being aware of what to do in the event of a gas leak [3]. As a result, a well-built forecasting model is essential for managing energy policy successfully by offering energy diversity and requirements that adapt to the dynamic structure of a country, region, or the world, in line with unprecedented increases in energy demand [4].

Natural gas is a type of fossil fuel [5]. Utilizing natural gas to meet energy needs for businesses, transportation, and other uses is environmentally not harmful and legally accepted [6]. Energy usage in buildings is a crucial factor in global efforts to achieve energy sustainability. According to [7], buildings, including both residential and commercial sectors, account for a fifth to a third of all energy use globally. This considerable share highlights the importance of advancing sustainable energy management practices within the built environment [8]. As urbanization continues to accelerate rapidly and energy demand increases, enhancing energy efficiency in buildings has become a strategic priority for achieving global climate targets, such as those

*Corresponding Author
Email: majuji@gsu.edu.ng

outlined in the United Nations Sustainable Development Goals (SDG 7: Affordable and Clean Energy) and various net-zero initiatives [9][10].

Currently, more than a third of the energy utilized in residential buildings in Europe comes from natural gas [11]. The strategic growth of nations' economies and societies depends on energy [12]. The primary purpose of data mining is to create models using preprocessed or existing data to identify patterns that the dataset's attributes have revealed [13]. Some of the patterns serve as explanations by illuminating the relationships between the qualities. Other modes are predictive, inferring potential values for particular attributes based on the most recent data [14]. Data mining is typically used to identify four distinct types of patterns: association patterns, prediction patterns, clustering patterns, and sequencing patterns [15]. Predicting the demand and consumption of natural gas for residential addresses using a hybrid method for efficiency is of paramount importance, as it will aid in planning and scheduling activities, including informed decision-making. In some countries, the demand for natural gas varies with the seasons; for instance, in Algeria, demand is higher during holidays and special days [3][16]. Likewise, in Nigeria, the demand for natural gas, particularly compressed natural gas, is typically higher during the last quarter of the year.

Building forecasting models to predict the quantity of a product that consumers will buy is known as demand forecasting [17]. It is influenced by a wide range of parameters, including the region, the number of customers, and the type of goods, among others [18]. Machine learning algorithms have gradually become popular tools for natural gas price forecasting [19]. In this era, an excellent predicting tool is machine learning algorithms and artificial intelligence.

Therefore, this study aims to design a machine learning model based on a data mining approach that can accurately and precisely predict natural gas demand and consumption at residential addresses [20]. Accurately predicting the demand and consumption of natural gas in both residential and commercial buildings would help reduce the impact of climate change, improve people's living conditions, and influence proper and appropriate private and government policies [21].

Machine learning methodology is the most widely used method that can be applied to both individual and group data. This study expands upon the definitions, models, parameters, and algorithms of machine learning, while also emphasizing the unique function of the machine learning approach in data mining [22]. As a result, we focus on the analysis and debates surrounding the problem of data mining using a learning method and provide a comparative study of the classification algorithms employed in data mining. Natural gas is in high demand due to its numerous benefits, and some models have been developed to predict natural gas usage [23].

The generality of the proposed model in testing datasets is the primary goal of building a prediction model using machine learning techniques [24]. Machine learning models can perform well on real data, despite a limited published framework for similar studies. The majority of the author's published studies are computationally expensive and require highly qualified forecasters to produce accurate and trustworthy results. In this work, we aim to enhance the accuracy of predicting natural gas demand and consumption in residential and commercial buildings.

The remainder of this paper is organized as follows: Section 2 presents the methodology; Section 3 consists of the results and discussion, and lastly, conclusions and limitations are drawn in Section 5.

2. METHOD

Machine learning is divided into traditional machine learning and deep learning models. Still, for this study, the earlier methods were used for Natural gas demand prediction for residential and commercial purposes, including the K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Decision Tree Regression (DTR), and Linear Regression (LR), as well as combined and hybrid methods [25]. Some of these models have been studied for several years, and their procedures have been established and refined. They focus not only on the prediction effect but also on practical problems, e.g., prediction effectiveness and data organization [26].

An Artificial Neural Network (ANN) was designed to simulate humans. It consists of an input layer, several hidden layers, and an output layer [27]. The units in adjacent layers are fully connected. It contains a considerable number of units and can theoretically approximate arbitrary functions; hence, it has a strong fitting ability, especially for nonlinear functions. Due to the complex model structure, training ANNs is time-consuming [28]. Remarkably, its models are trained using the backpropagation algorithm, which is not typically used to train deep networks. Thus, it belongs to machine learning models as seen in the above figure.

The Support Vector Machine (SVM) algorithm was used to find a max-margin separation hyperplane in the n-dimensional feature space [29][30]. It can be used to achieve gratifying results even with small-scale training sets, as the separation hyperplane is determined by only a small number of support vectors. However, SVMs are sensitive to noise near the decision boundary, also known as the hyperplane. SVMs can solve linear problems well. For nonlinear data, kernel functions are usually used. A kernel function maps the original space into a new space so that the original nonlinear data can be separated. The Core of SVM tricks is widely applied among both SVMs and other machine learning algorithms for prediction [31].

K-nearest neighbor (KNN) was designed to depend on the locality assumption, but can also support the manifold assumption when modeling high-dimensional data structures [32]. If most of a sample's neighbors belong to the same class, the sample has a high probability of belonging to the class. Thus, the classification result is only related to the top-k nearest neighbors. The parameter k greatly influences the performance of KNN models. The smaller k is, the more complex the model is, and the higher the risk of overfitting. Conversely, the larger k is, the simpler the model is and the weaker the fitting ability [33].

The Naïve Bayes machine learning algorithm is based on the conditional probability and the hypothesis of attribute independence. In each sample, the Naïve Bayes classifier calculates the conditional probabilities for different classes. The sample is classified into the class with the maximum likelihood [34]. The conditional probability formula is calculated as follows.

$$(X = x|Y = c_k) = \prod_{i=1}^{n} P(X^{(i)} = x^{(i)}|Y = c_k) \quad (1)$$

If the attribute independence hypothesis is satisfied, the Naïve Bayes algorithm reaches the optimal result. Unfortunately, that hypothesis is difficult to satisfy in reality; hence, the Naïve Bayes algorithm does not perform well on attribute-related data [21] [35].

The logistic regression (LR) is a type of log-linear model [22][36]. It computes the probabilities of different classes through a parametric logistic distribution, calculated as shown below.

$$P(Y = k|x) = \frac{e^{w_{k^*x}}}{1 + \sum_{k}^{k-1} e^{w_{k^*x}}} \quad (2)$$

Where k = 1, 2...K − 1. The sample x is classified into the class with the maximum probability. An LR model is easy to construct, and model training is efficient. However, LR struggles to handle nonlinear data, which limits its application.

The decision tree algorithm classifies data using a series of rules. The model is tree-like, which makes it interpretable. It can automatically exclude irrelevant and redundant features in the sample dataset. The learning process includes feature selection, tree generation, and tree pruning. When training a decision tree model, the algorithm selects the most suitable features individually and generates child nodes from the root node. The decision tree is one of the basic machine learning classifiers. The advanced algorithms, such as the random forest and the extreme gradient boosting (XGBoost), consist of multiple decision trees [37].

In clustering, data points are grouped based on a defined measure of similarity or distance, with the assumption that points with similar characteristics belong to the same cluster [38]. This approach involves grouping highly identical data into the same clusters and less similar data into different clusters. Unlike classification, clustering is a type of unsupervised learning. No prior knowledge or labeled data is required for clustering algorithms; hence, the dataset requirements are relatively low [39]. However, when using clustering algorithms to detect attacks, it is necessary to refer to external information.

K-means is a typical clustering algorithm, where K is the number of clusters and the means are the means of attributes. The K-means algorithm uses distance as a standard measure of similarity. The smaller the distance between two data objects is, the more likely they are to be placed in the same cluster. The K-means algorithm performs well on linear data, but its results on non-linear data are not ideal. In addition, the K-means algorithm is sensitive to the initialization condition and the parameter K. Consequently, numerous repeated experiments must be conducted to determine the appropriate parameter value [40].

All machine learning algorithms have individual strengths and weaknesses. A natural approach is to combine various weak classifiers to implement a strong classifier. Ensemble methods train multiple classifiers, which vote to obtain the final results. Hybrid methods are designed with multiple stages, each of which utilizes a classification model. Because ensemble and hybrid classifiers typically outperform single classifiers, an increasing number of researchers have begun to investigate these types of classifiers. The key points lie in selecting which classifiers to combine and how they are combined [41].

It is an unrefined fuel. Natural gas is better suited to reducing environmental pollution than refined fuel oils, such as gasoline and diesel. Natural gas is a type of fossil fuel. Utilizing natural gas to meet energy needs for businesses, transportation, and other uses is environmentally benign. Buildings account for a fifth to a third of all energy use worldwide. Currently, more than a third of the energy utilized in residential buildings in Europe comes from natural gas [6][42]. The strategic growth of nations' economies and societies depends on energy. Predicting energy consumption has become a significant challenge in many countries, driven by the exponential increase in global energy demand over the past few decades. About 34.7% of the world's energy is consumed by the residential and commercial sectors [24]. The use of natural gas has been proposed as a means to enhance energy supply security and reduce global environmental pollution [11][43].

It is an approach that combines techniques from machine learning, statistics, and database systems to extract and identify patterns in massive datasets. The use of data mining techniques in natural gas systems for daily load forecasting is becoming increasingly common. The primary purpose of data mining is to create models using preprocessed or existing data to identify patterns that the dataset's attributes have revealed. Some of the patterns serve as explanations by illuminating the relationships between the qualities. Other modes are predictive, inferring potential values for particular attributes based on the most recent data. Data mining is typically used to identify four distinct types of patterns: association patterns, prediction patterns, clustering patterns, and sequencing patterns [44].

There is a long history of ensemble learning techniques outperforming other machine learning methods. Classification and regression issues are among the application domains of these programs. Well-known ensemble models that combine weak learners to form an ensemble include the random forest model and the gradient boosting model [45]. These models have a homogeneous collection of weak learners, which means that weak learners of the same type are grouped to demonstrate their combined strength. In this study, we created a hybrid ensemble learning model using a diverse group of weak learners. In this challenge, various machine learning methods are combined to tackle a regression problem.

A machine learning concept called ensemble learning leverages the combined power of machine learning models to solve learning problems for both classification and regression tasks. This method classifies many homogeneous machine learning models as weak learners and groups them. Each of the weak learners exhibits its unique result when applied to the problem, whether it be on the complete training set or only a portion of it [46]. The ultimate result is obtained by combining the findings of each poor learner.

There are two well-known types of ensemble learning: bagging and boosting. The renowned ensemble learning model for bagging is known as the random forest. Another well-liked ensemble learning approach that falls under the boosting umbrella is AdaBoost [47]. While the boosting models use the complete dataset, the bagging models only use a portion of it.

In this research work, a hybrid ensemble learning model is created using four types of machine learning models as weak learners. These models include K-Nearest Neighbor Regression, Support Vector Machine Regression, Decision Tree Regression, and Linear Regression. While a homogeneous group of weak learners is utilized for some experiments, this work employed a heterogeneous group of weak learners, hence the term 'hybrid'. Figure 3 shows the architecture of the proposed approach. The regression job for predicting natural gas demand modeling was applied using a hybrid ensemble learning model created from four weak learning models.

3. RESULTS AND DISCUSSION

A dataset is a fundamental step in the data analysis process, as it provides a contextual and critical understanding of the dataset before more complex analyses are performed [48]. It helps uncover data quality issues, explore relationships between variables, and generate hypotheses for further investigation. As proposed in the methodology, we first examine the historical nature of natural gas demand values from the dataset before applying a variety of machine learning algorithms to make forecasts. We evaluate the effectiveness of machine learning models by contrasting estimated values with actual consumption levels after using the models. The most extended period dataset in the literature was collected from the Kaggle machine learning repository.

Model building typically involves several steps, including selecting an appropriate algorithm, defining the model structure and parameters, training the model on a dataset, and evaluating its performance. The choice of algorithm and model structure will depend on the nature of the problem being addressed, as well as the available data and computational resources [49]. For this study, decision trees (DTs), support vector machines (SVMs), and hybrid ensemble models were employed. Train-test split is a widely used technique in machine learning and is an essential step in model evaluation and selection. It helps to prevent over-fitting, where the model performs well on the training data but poorly on new data, by evaluating the model's performance on a separate testing set. Typically, the training set is much larger than the testing set, as in this study, with a common split being 75% training data and 25% testing data. However, the optimal split will depend on the nature of the problem and the available data.

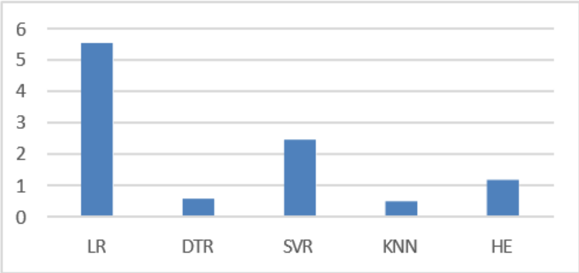

Figure 1. Mean Squared Error (MSE)

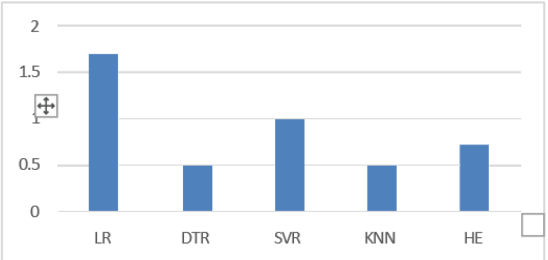

Figure 2. Mean Absolute Error (MAE)

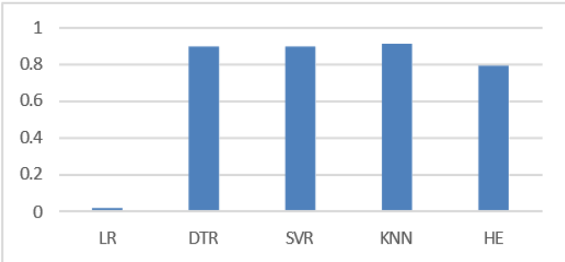

Figure 3. R^2

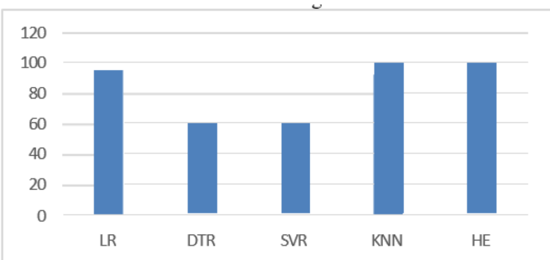

Figure 4. Accuracy Score

Figure 1 presents the mean square error (MSE), also known as Root Mean Square Error (RMSE). It is similar to the mean absolute error in that it provides a rough idea of the magnitude of the error. The bar chart in Figure 1 illustrates a comparative evaluation of five machine learning models, Linear Regression (LR), Decision Tree Regression (DTR), Support Vector Regression (SVR), K-Nearest Neighbors (KNN), and a Hybrid Ensemble (HE), across dissimilar performance metrics, providing a holistic view of their strengths and limitations.

In Figure 1, LR records the highest value (~6), followed by SVR (~2.5) and HE (~1.5), while DTR and KNN exhibit the lowest values (~0.5). Assuming the metric represents an error measure (e.g., RMSE or MAE), this suggests that DTR and KNN perform best. In contrast, LR performs poorly, likely due to its limitation in modeling non-linear relationships.

Figure 2 shows a notable improvement in LR's score (~1.6), but it still lags behind other models. SVR (~1.0) and HE (~0.7) offer intermediate performance, while DTR and KNN (~0.5) again emerge as top performers, reinforcing their reliability for this task. The mean absolute error (MAE) is shown in. Figure 2 above illustrates the extent of the incorrect predictions. As in the MSE, the result of MAE gives similar performance results for the various machine learning models.

As shown in Figure 3, the results of the R-squared evaluation model for all models are presented. It indicates the goodness of fit of a set of predictions to the actual values. It is used to verify and confirm that the prediction performance results obtained from the designed model are accurate.

Again, in Figure 3, the metric has shifted to a positive performance indicator, where DTR, SVR, and KNN all achieve high values (~0.9), demonstrating excellent predictive accuracy. The HE models also perform well (~0.8), though slightly lower than the individual non-linear models. In stark contrast, LR shows a near-zero score, indicating that it fails to explain the variance in the data and is ill-suited for the problem at hand.

The performance metrics for accuracy are shown in Figure 4, which indicates that the hybrid ensemble machine learning model achieves the highest accuracy, followed by linear regression, and then K-nearest neighbor.

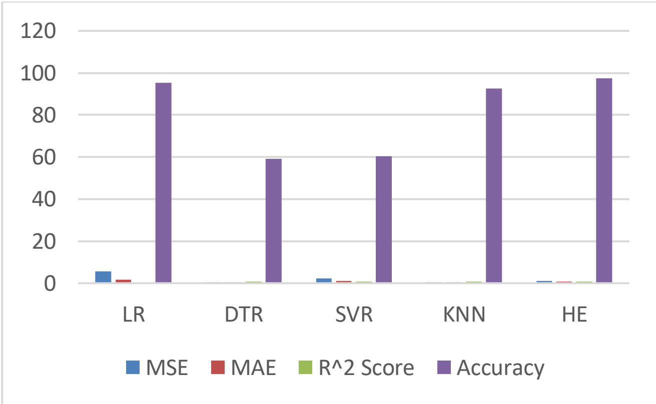

Figure 5. Comparison of Models Used

Figure 5 shows the comparison of all the various machine learning models used in the study, as also shown in Table 1 below. As displayed in the figure above the hybrid ensemble has the highest prediction accuracy and minimum performance error. Therefore, for efficient energy management and decision-making, a precise forecast of residential natural gas demand is essential. For energy management organizations and regulators, the hybrid ensemble regression approach provides a powerful tool for making informed decisions about resource allocation, infrastructure development, and energy-saving strategies. By ensuring reliable and efficient natural gas transportation, energy companies can streamline their supply chains, reduce waste, and enhance customer satisfaction.

Table 1. Comparison of Prediction Models' Performance

MODEL	MSE	MAE	R^2	Accuracy
Linear Regression	5.5179	1.6928	0.0174	95.3845
Decision Tree Regressor	0.5705	0.4872	0.8984	59.0136
Support Vector Regressor	2.4484	0.9959	0.8984	60.4937
K-Nearest Neighbor	0.4894	0.4440	0.9128	92.4553
Hybrid-Ensemble model	1.1648	0.7216	0.7925	97.4871

To predict residential natural gas consumption, the use of a hybrid ensemble regression machine learning approach offers several noteworthy benefits and consequences, as noted in [35][50]. This discussion examines the approach's main principles and any potential field repercussions. When compared to individual regression models, the hybrid ensemble regression approach showed greater prediction accuracy [36][51]. The method captures a broader range of patterns and correlations contained in the data by merging different algorithms, such as LR, SVR, DTR, and KNN. As a result, estimates of residential natural gas demand become more precise and trustworthy, which is very advantageous for energy management firms, decision-makers, and energy providers when planning and maximizing resource allocation.

The use of ensemble techniques improves the prediction model's resilience and generalization capabilities. The hybrid ensemble strategy reduces the likelihood of overfitting and enhances model stability by incorporating a range of regression models. This is crucial when predicting residential natural gas demand because the addition of numerous variables, such as socioeconomic and meteorological variables, might result in complex interactions. The hybrid ensemble approach makes sure the model can handle various situations and adjust to changes in the data.

The hybrid ensemble regression approach is versatile and can be tailored to a wide range of prediction scenarios. The ensemble process's fusion techniques, including stacking or weighted averaging, enable various combinations of individual models. Furthermore, the hybrid ensemble approach can be easily expanded to include additional regression models or domain-specific knowledge, thereby further increasing its flexibility.

4. CONCLUSION AND LIMITATION

The study concludes that the hybrid ensemble regression machine learning methodology is a significant technique for accurately anticipating residential natural gas demand and consumption. By combining multiple regression models, the proposed hybrid ensemble method enhances predictive performance while capturing a variety of patterns and relationships within the data. The individual model, K-Nearest Neighbor (KNN) delivered the best overall predictive performance with an R^2 of 0.9128, MAE of 0.4440, and accuracy of 92.46%, outperforming both Decision Tree Regressor and Support Vector Regressor. While the Hybrid Ensemble (HE) model achieved slightly lower R^2 (0.7925), it demonstrated the highest accuracy (97.49%), signifying its potential to capture generalized patterns across diverse conditions. In contrast, Linear Regression exhibited poor predictive strength (R^2 = 0.0174) despite a deceptively high accuracy, highlighting the limitations of relying solely on linear assumptions or accuracy as a standalone metric. These results highlight the importance of utilizing non-linear and ensemble learning methods to enhance the precision and robustness of energy consumption forecasting, a crucial step toward achieving more efficient and sustainable energy systems in the building sector. Furthermore, more similar studies and the use of the hybrid ensemble approach can advance the topic and help achieve sustainable energy management goals in smart city infrastructure, as well as efficiency goals within industrial manufacturing systems.

REFERENCES

[1] Y. Li, Y. He, & M. Zhang, "Prediction of Chinese Energy Structure Based on Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM)", *Energy Science & Engineering*, vol. 8, no. 8, pp. 2680-2689, 2020. https://doi.org/10.1002/ese3.698

[2] S. Ahmed, S. Madanian, and F. Mirza, "Prediction of Natural Gas Consumption in Bahçe ş ehir Using Machine Learning Prediction of Natural Gas Consumption in Bahçe ş ehir Using Machine Learning Models," *ACIS 2020 Proceedings*, 2021. https://aisel.aisnet.org/acis2020/12

[3] O. Laib, M. Khadir, & L. Mihaylova, "Toward Efficient Energy Systems Based on Natural Gas Consumption Prediction with LSTM Recurrent Neural Networks", *Energy*, vol. 177, pp. 530-542, 2019. https://doi.org/10.1016/j.energy.2019.04.075

[4] L. Chen and J. Zhang, "A Forecast Model of City Natural Gas Daily Load Based on Data Mining", *Scientific Programming*, vol. 2022, pp. 1-14, 2022. https://doi.org/10.1155/2022/1562544

[5] Ö. Beyca, B. Ervural, E. Tatoğlu, P. Özuyar, & S. Zaim, "Using Machine Learning Tools for Forecasting Natural Gas Consumption in the Province of Istanbul", *Energy Economics*, vol. 80, pp. 937-949, 2019. https://doi.org/10.1016/j.eneco.2019.03.006

[6] D. Then, C. Spalthoff, J. Bauer, T. Kneiske, & M. Braun, "Impact of Natural Gas Distribution Network Structure and Operator Strategies on Grid Economy in Face of Decreasing Demand", *Energies*, vol. 13, no. 3, pp. 664, 2020. https://doi.org/10.3390/en13030664

[7] N. Tian, B. Shao, G. Bian, H. Zeng, X. Li, & W. Zhao, "Application of Forecasting Strategies and Techniques to Natural Gas Consumption: A Comprehensive Review and Comparative Study", *Engineering Applications of Artificial Intelligence*, vol. 129, pp. 107644, 2024. https://doi.org/10.1016/j.engappai.2023.107644

[8] T. Le, M. Vo, B. Vo, E. Hwang, S. Rho, & S. Baik, "Improving Electric Energy Consumption Prediction Using CNN and Bi-LSTM", *Applied Sciences*, vol. 9, no. 20, pp. 4237, 2019. https://doi.org/10.3390/app9204237

[9] A. Mehbodniya, I. Alam, S. Pande, R. Neware, K. Rane, M. Shabaz et al., "Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques", *Security and Communication Networks*, vol. 2021, pp. 1-8, 2021. https://doi.org/10.1155/2021/9293877

[10] M. S. U. Miah, J. Sulaiman, M. I. Islam, M. Masuduzzaman et al., "Predicting Short-Term Energy Demand in Smart Grid: A Deep Learning Approach for Integrating Renewable Energy Sources in Line with SDGs 7, 9, and 13," *ResearchGate Preprint*, 2023. https://doi.org/10.48550/arXiv.2304.03997

[11] A. Hajikhani and A. Suominen, "Mapping the Sustainable Development Goals (SDGs) in Science, Technology and Innovation: Application of Machine Learning in SDG-Oriented Artefact Detection", *Scientometrics*, vol. 127, no. 11, pp. 6661-6693, 2022. https://doi.org/10.1007/s11192-022-04358-x

[12] F. Shaikh and D. Rao, "Prediction of Cancer Disease using Machine Learning Approach", *Materials Today: Proceedings*, vol. 50, pp. 40-47, 2022. https://doi.org/10.1016/j.matpr.2021.03.625

[13] M. Su, Z. Zhang, Y. Zhu, D. Zha, & W. Wen, "Data Driven Natural Gas Spot Price Prediction Models Using Machine Learning Methods", *Energies*, vol. 12, no. 9, pp. 1680, 2019. https://doi.org/10.3390/en12091680

[14] K. Yin and S. Htay, "Prediction of Natural Gas Final Consumption using Artificial Neural Networks", *2020 International Conference on Advanced Information Technologies (ICAIT)*, pp. 224-229, 2020. https://doi.org/10.1109/icait51105.2020.9261813

[15] X. Liu, G. Hou, & L. Yang, "Dynamic Combined Optimal Scheduling of Electric Energy and Natural Gas Energy Consumption in Data Center", *Discrete Dynamics in Nature and Society*, vol. 2022, no. 1, 2022. https://doi.org/10.1155/2022/3917170

[16] A. Abubakar, M. Ajuji, & I. Yahya, "DeepFMD: Computational Analysis for Malaria Detection in Blood-Smear Images Using Deep-Learning Features", *Applied System Innovation*, vol. 4, no. 4, pp. 82, 2021. https://doi.org/10.3390/asi4040082

[17] D. Šebalj, J. Mesarić, & D. Dujak, "Analysis of Methods and Techniques for Prediction of Natural Gas Consumption", *Journal of Information and Organizational Sciences*, vol. 43, no. 1, pp. 99-117, 2019. https://doi.org/10.31341/jios.43.1.6

[18] M. Sharma, B. Raj, B. Ramamurthy, & R. Bhaskar, "Credit Card Fraud Detection Using Deep Learning Based on Auto-Encoder", *ITM Web of Conferences*, vol. 50, pp. 01001, 2022. https://doi.org/10.1051/itmconf/20225001001

[19] M. Shapi, N. Ramli, & L. Awalin, "Energy Consumption Prediction by Using Machine Learning for Smart Building: Case Study in Malaysia", *Developments in the Built Environment*, vol. 5, pp. 100037, 2021. https://doi.org/10.1016/j.dibe.2020.100037

[20] F. Gao and X. Shao, "Forecasting Annual Natural Gas Consumption via the Application of a Novel Hybrid Model", *Environmental Science and Pollution Research*, vol. 28, no. 17, pp. 21411-21424, 2021. https://doi.org/10.1007/s11356-020-12275-w

[21] L. Lok, V. Hameed, & M. Rana, "Hybrid Machine Learning Approach for Anomaly Detection", *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 2, pp. 1016, 2022. https://doi.org/10.11591/ijeecs.v27.i2.pp1016-1024

[22] J. Jouhar, A. Pratap, N. Tijo, & M. Mony, "Fake News Detection using Python and Machine Learning", *Procedia Computer Science*, vol. 233, pp. 763-771, 2024. https://doi.org/10.1016/j.procs.2024.03.265

[23] J. Li, Y. Guo, X. Zhang, & Z. Fu, "Using Hybrid Machine Learning Methods to Predict and Improve the Energy Consumption Efficiency in Oil and Gas Fields", *Mobile Information Systems*, vol. 2021, pp. 1-7, 2021. https://doi.org/10.1155/2021/5729630

[24] S. Nabavi, A. Aslani, M. Zaidan, M. Zandi, S. Mohammadi, & N. Motlagh, "Machine Learning Modeling for Energy Consumption of Residential and Commercial Sectors", *Energies*, vol. 13, no. 19, pp. 5171, 2020. https://doi.org/10.3390/en13195171

[25] T. Srinivas, A. Madhusudhan, J. Dhanraj, R. Sekaran, N. Mostafaeipour, N. Mostafaeipour et al., "Novel Based Ensemble Machine Learning Classifiers for Detecting Breast Cancer", *Mathematical Problems in Engineering*, vol. 2022, pp. 1-16, 2022. https://doi.org/10.1155/2022/9619102

[26] S. Yang, J. Wu, Y. Du, Y. He, & X. Chen, "Ensemble Learning for Short-Term Traffic Prediction Based on Gradient Boosting Machine", *Journal of Sensors*, vol. 2017, pp. 1-15, 2017. https://doi.org/10.1155/2017/7074143

[27] Y. Hendrawan and O. Chekuri, "Exploring Supervised Learning Methods for Predicting Cuisines from Their Ingredients", *Vokasi Unesa Bulletin of Engineering, Technology and Applied Science*, vol. 2, no. 1, pp. 67-76, 2025. https://doi.org/10.26740/vubeta.v2i1.34153

[28] D. Batra, N. Diwan, U. Upadhyay, J. Kalra, T. Sharma, A. Sharma et al., "RecipeDB: A Resource for Exploring Recipes", *Database*, vol. 2020, 2020. https://doi.org/10.1093/database/baaa077

[29] M. Atzeni, S. Dhuliawala, K. Murugesan, and M. Sachan, "Case-based Reasoning for Better Generalization in Textual Reinforcement Learning," *Arxiv Preprint*, 2021. https://doi.org/10.48550/arXiv.2110.08470

[30] E. Y. Hamedani and N. S. Aybat, "A Primal-Dual Algorithm with Line Search for General Convex-Concave Saddle Point Problems," *Arxiv Preprint*, vol. 31, no. 2, pp. 1299–1329, 2021. https://doi.org/10.48550/arXiv.1803.01401

[31] A. Abubakar, M. Ajuji, & I. Yahya, "Comparison of Deep Transfer Learning Techniques in Human Skin Burns Discrimination", *Applied System Innovation*, vol. 3, no. 2, pp. 20, 2020. https://doi.org/10.3390/asi3020020

[32] A. Abubakar and H. Ugail, "Discrimination of Human Skin Burns Using Machine Learning", *Advances in Intelligent Systems and Computing*, pp. 641-647, 2019. https://doi.org/10.1007/978-3-030-22871-2_43

[33] M. Ajuji, A. A. Lamido, M. Jungudo, and A. J. Kawu, "Hybrid Ensemble Machine Learning Approach for Cancer Prediction," *BIMA Journal of Science and Technology*, vol. 8, no. 1b, pp. 206–217, 2024. https://doi.org/10.56892/bima.v8i1.627

[34] M. Applalanaidu and G. Kumaravelan, "A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification", *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, pp. 716-724, 2021. https://doi.org/10.1109/icicv50876.2021.9388488

[35] P. Trivedi, Y. Narayan, V. Ravi, P. Kumar, P. Kaur, K. Tabianan et al., "Plant Leaf Disease Detection and Classification Using Segmentation Encoder Techniques", *The Open Agriculture Journal*, vol. 18, no. 1, 2024. https://doi.org/10.2174/0118743315321139240627092707

[36] D. Caballero-Ramirez, Y. Báez-López, J. Limón-Romero, G. Tortorella, & D. Tlapa, "An Assessment of Human Inspection and Deep Learning for Defect Identification in Floral Wreaths", *Horticulturae*, vol. 9, no. 11, pp. 1213, 2023. https://doi.org/10.3390/horticulturae9111213

[37] C. Cevallos, H. Pönce, E. Moya-Albor, & J. Brieva, "Vision-Based Analysis on Leaves of Tomato Crops for Classifying Nutrient Deficiency using Convolutional Neural Networks", *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-7, 2020. https://doi.org/10.1109/ijcnn48605.2020.9207615

[38] J. Pan, X. Qiao, C. Zhang, B. Li, L. Li, G. Li et al., "Stacking Ensemble Learning-Assisted Simulation of Plasma-Catalyzed CO2 Reforming of Methane", *Electronics*, vol. 14, no. 7, pp. 1329, 2025. https://doi.org/10.3390/electronics14071329

.

[39] B. Shi, W. Xiao, L. Zhang, T. Wang, Y. Jiang, J. Shang et al., "Multi-Objective Optimization Method for Power Transformer Design Based on Surrogate Modeling and Hybrid Heuristic Algorithm", *Electronics*, vol. 14, no. 6, pp. 1198, 2025. https://doi.org/10.3390/electronics14061198
[40] S. Tayebati and K. Cho, "Machine Learning Framework for Hybrid Clad Characteristics Modeling in Metal Additive Manufacturing", *Journal of Manufacturing and Materials Processing*, vol. 9, no. 2, pp. 49, 2025. https://doi.org/10.3390/jmmp9020049
[41] A. Mohsenimanesh and E. Entchev, "Charging Strategies for Electric Vehicles Using a Machine Learning Load Forecasting Approach for Residential Buildings in Canada", *Applied Sciences*, vol. 14, no. 23, pp. 11389, 2024. https://doi.org/10.3390/app142311389
[42] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation", *Neural Computation*, vol. 15, no. 6, pp. 1373-1396, 2003. https://doi.org/10.1162/089976603321780317
[43] A. Mostafa and H. Mahmoud, "Review of Data Mining Concept and its Techniques", *International Journal of Academic Research in Business and Social Sciences*, vol. 12, no. 6, 2022. https://doi.org/10.6007/ijarbss/v12-i6/13135
[44] D. Li, Z. Qi, Y. Zhou, & M. Elchalakani, "Machine Learning Applications in Building Energy Systems: Review and Prospects", *Buildings*, vol. 15, no. 4, pp. 648, 2025. https://doi.org/10.3390/buildings15040648
[45] J. Ji, H. Yu, X. Wang, & X. Xu, "Machine Learning Application in Building Energy Consumption Prediction: A Comprehensive Review", *Journal of Building Engineering*, vol. 104, pp. 112295, 2025. https://doi.org/10.1016/j.jobe.2025.112295
[46] I. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects", *IEEE Access*, vol. 10, pp. 99129-99149, 2022. https://doi.org/10.1109/access.2022.3207287
[47] J. Liu, S. Wang, N. Wei, X. Chen, H. Xie, & J. Wang, "Natural Gas Consumption Forecasting: A Discussion on Forecasting History and Future Challenges", *Journal of Natural Gas Science and Engineering*, vol. 90, pp. 103930, 2021. https://doi.org/10.1016/j.jngse.2021.103930
[48] H. Sayegh, A. Leconte, G. Fraisse, É. Wurtz, & S. Rouchier, "Multi Objective Optimization of Detailed Building Models with Typical Short Sequences Considering Sequential and Adaptive Methods", *Engineering Applications of Artificial Intelligence*, vol. 118, pp. 105645, 2023. https://doi.org/10.1016/j.engappai.2022.105645
[49] H. Yang, M. Ran, H. Feng, & D. Hou, "K-PCD: A New Clustering Algorithm for Building Energy Consumption Time Series Analysis and Predicting Model Accuracy Improvement", *Applied Energy*, vol. 377, pp. 124584, 2025. https://doi.org/10.1016/j.apenergy.2024.124584
[50] Y. Qi, W. Hu, Y. Dong, Y. Fan, L. Dong, & M. Xiao, "Optimal Configuration of Concentrating Solar Power in Multienergy Power Systems with An Improved Variational Autoencoder", *Applied Energy*, vol. 274, pp. 115124, 2020. https://doi.org/10.1016/j.apenergy.2020.115124
[51] D. Gao, Y. Zhi, X. Rong, & X. Yang, "Mismatch Analysis of Rooftop Photovoltaics Supply and Farmhouse Load: Data Dimensionality Reduction and Explicable Load Pattern Mining via Hybrid Deep Learning", *Applied Energy*, vol. 377, pp. 124520, 2025. https://doi.org/10.1016/j.apenergy.2024.124520

BIOGRAPHIES OF AUTHORS

Mohammed Ajuji is a lecturer in the Department of Computer Science, Faculty of Science, Gombe State University, Gombe, Nigeria. He received a Bachelor of Science (BSc) from the Gombe State University (GSU) in Computer Science. He also received a Master of Science (MSc) from the Abubakar Tafawa Balewa University (ATBU), Bauchi, Nigeria in 2023. He has published papers mainly in the fields of machine learning, artificial intelligence, and computer vision. He is currently undergoing his PhD program at the Modibbo Adama University, Yola. He can be contacted at email: majuji@gsu.edu.ng.

Muhammad Dawaki is a lecturer in the Department of Computer Science, at Gombe State University, Gombe, Nigeria. He received a Bachelor of Science (BSc) from the Gombe State University (GSU) in Computer Science, in 2015. He also received a Master of Science (MSc) from the University of Leeds in 2018. He is mainly researching data mining, artificial intelligence, and blockchain technology. He can be contacted at email: dawaki@gsu.edu.ng.