A REVIEW OF THE MACHINE LEARNING APPROACH FOR PREDICTING RESIDENTIAL NATURAL GAS DEMAND AND CONSUMPTION

AMINU, A.[1]*, UMAR, F.Z.[2], YAU, A.G.[2], AJUJI, M.[3], PYELSHAK, Y.[4] AND OKONKWO, O.G.[5]

[1]Department of Computer Sciences, Nigerian Army University, Biu, Nigeria
[2]Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi, Nigeria
[3]Department of Computer Science, Gombe State University, Gombe, Nigeria
[4]Department of Health Information Management, Plateau State College of Health, Zawan, Nigeria
[5]Department of Computer Sciences, Federal Polytechnic, Bauchi, Nigeria

ABSTRACT

The only byproducts of burning natural gas are carbon dioxide, water vapor, and very little amounts of nitrogen oxide, making it the cleanest fossil fuel in the world. Natural gas is also used to power a large variety of consumer goods, including furnaces, dryers, fireplaces, and stoves. You definitely use natural gas to power at least one of your appliances. This study provides a complete evaluation of the machine learning forecasting approaches currently in use for estimating residential natural gas demand and consumption. The paper discusses the main difficulties in estimating natural gas demand, including the complicated and dynamic interaction between weather patterns and energy consumption as well as the diversity of homes and their energy consumption habits. The paper also provides an overview of the various machine learning methods and algorithms, such as artificial neural networks, decision tree models, and regression-based approaches that are utilized for extrapolating natural gas demand. The research examines these methods' precision, memory, and accuracy in addition to their benefits and drawbacks. In its last section, the report makes several recommendations for more research, including the use of data from smart meters and the implementation of advanced machine learning techniques, such as deep learning, to improve forecast accuracy. The state-of-the-art in machine learning for natural gas prediction and consumption is generally covered in this review, which is an essential resource for academics and business experts.

Keywords: Demand, Consumption, Energy, Machine learning and Natural Gas
***Correspondence:** aminuagabus@gmail.com, 07032132336

INTRODUCTION

Natural gas has been employed in all parts of human life since the beginning of time, particularly in home construction. Because it only produces carbon dioxide, water vapor, and very little amounts of nitrogen oxide when burned, natural gas is the cleanest fossil fuel on the planet. Stoves, dryers, fireplaces, and furnaces are just a few of the many consumer products that are powered by natural gas. Most likely, at least one of your appliances runs on natural gas. Almost all energy sources, including natural gas, can be dangerous if used improperly. By following a few fundamental safety guidelines and being informed of what to do in the event of a gas leak, you can keep yourself and your loved ones safe. As a result, to properly manage energy policy, a forecasting model must be well-built and provide energy diversity as well as energy requirements that can be adjusted to the dynamic structure of a nation, region, or the entire world in line with previously unheard-of increases in energy demand [1].

As a fossil fuel, natural gas is one of them. It is environmentally friendly to use natural gas to supply energy requirements for industry, transportation, and other uses. Between a fifth and a third of the energy used worldwide is consumed by buildings. Currently, natural gas accounts for more than one-third of the energy used in both residential and commercial buildings in Europe [2]. Energy is necessary for the strategic development of civilizations and economies. The fundamental goal of data mining is to build models from preprocessed or existing data to discover patterns that the features of the data set have revealed. Some of the patterns provide explanations by showing how the attributes are related to one another. Inferring potential values for specific traits based on the most current data, other modes are predictive. Four types of patterns are commonly found via data mining: association patterns, prediction patterns, clustering patterns, and sequencing patterns [3].

Predicting the amount of natural gas demand and consumption needed for residential addresses is crucial because it will aid in organizing and scheduling tasks, as well as helping to make wise decisions. Natural gas consumption varies by season in some nations, such as Algeria, where it is higher on holidays and special days [4]. Similar to other countries, Nigeria has a heightened demand for natural gas, particularly premium motor spirit, during the final quarter of the year.

Demand forecasting is the practice of developing forecasting models to predict how much of a product a consumer will buy. It depends on some things, like the size of the warehouse, the number of customers, the type of goods, etc. Forecasting client demand gets significantly more challenging as there are more distribution sites and items available. Machine learning algorithms have steadily grown in favor of forecasting natural gas prices [5]. In the present era, machine learning algorithms and artificial intelligence are excellent prediction tools.

As a result, the goal of this work is to develop a machine learning model using a data mining approach that can accurately and precisely estimate natural gas demand and ingesting at residential addresses. The impact of climate change may be lessened, people's living conditions could be improved, and correct and appropriate private and governmental policies could be influenced by accurately forecasting the demand for and consumption of regular or natural gas in both residential and commercial structures [6].

Data mining is the most popular technique in the data science industry and a machine learning methodology that can map into both individual and group data. This study emphasizes the special role that the machine learning method plays in data mining while also extending the definitions, models, parameters, and algorithms of machine learning. In order to provide a comparative analysis of the classification algorithms used in data mining, we focus on the analysis and discussions of the problem of data mining using a learning technique. Due to its many advantages, natural gas is in high demand, and several models have been developed to predict natural gas use [7].

This study focuses on traditional machine learning algorithms, such as the support vector machine (SVM), linear regression (LR), boosting algorithms (adaptive boost (AdaBoost), extreme gradient boost (XGBoost), Decision Tree (DT), and others, for predicting natural gas demand and consumption in residential and commercial buildings. The following is how this article is divided into its remaining sections: Sections 2 and 3 provide an overview of a few machine learning prediction techniques that have been discussed in the literature, Section 4 presents machine learning applications in energy prediction, Section 5 examines the research gaps that still need to be filled and offers research recommendations for future work, and Section 6 presents a framework to increase the accuracy and reliability of natural disaster predictions.

Advantage and inspiration

Machine learning prediction has several benefits, some of which are as follows [8, 9]: Machine learning algorithms may identify complex links and patterns in data that human analysts might overlook. Better forecasting and decision-making could result from this. Because they can efficiently and quickly analyze enormous amounts of data, machine learning algorithms are especially well suited for tasks requiring high-volume data processing. Machine learning techniques can be used to benefit a wide range of fields and applications, such as image and audio recognition, natural language processing, and predictive analytics.

The following are some examples of machine learning prediction model inspirations: enhanced decision-making: Machine learning prediction can offer insightful analyses of complicated data patterns that can aid in the development of better informed and precise decisions [10]. By enabling businesses to innovate and react swiftly to shifting market conditions, improved customer experiences and machine learning prediction can give them a competitive advantage.

General background of prediction techniques

Prediction techniques refer to a set of methods and algorithms used to forecast future events or outcomes based on historical data and patterns [11]. These techniques are used in various fields such as finance, economics, healthcare, and engineering to make predictions about future trends, behaviors, and events. There are several approaches to prediction techniques, including statistical methods, machine learning, and artificial intelligence [10, 12]. Statistical methods involve using mathematical models and probability theory to analyze historical data and make predictions about future outcomes. These methods include regression analysis, time series analysis, and forecasting techniques such as moving averages and exponential smoothing.

Machine learning is a subfield of artificial intelligence that deals with training computers to recognize patterns in data and predict the results of brand-new data. Machine learning techniques such as supervised learning, unsupervised learning, and reinforcement learning are extensively used in fields such as natural language processing, image identification, and predictive analytics. Artificial intelligence, which encompasses rule-based systems, fuzzy logic, and expert systems, is a far broader topic than machine learning alone [12]. Artificial intelligence (AI) techniques are used to create intelligent machines that are capable of understanding natural language, identifying images, and making decisions for all tasks that usually need human intelligence [13]. Prediction techniques are essential in many businesses because they enable us to base our decisions on historical information and patterns. The choice of prediction technique is influenced by the type of data, the complexity of the problem, and the required degree of accuracy and reliability of the forecasts [14].

Machine learning approaches

The prediction of natural gas demand in residential and commercial settings will be done in this study using traditional machine learning models, including artificial neural networks, support vector machines, K-nearest neighbor, nave Bayes, logistic regression, decision trees, clustering, and combined and hybrid methods [16]. The categories of deep learning models and conventional machine learning, on the other hand, can be further subdivided. Some of these models have been the subject of in-depth study, and their methodology has been improved. Along with the prediction effect, they also take into account practical concerns like data or organizational challenges and forecast efficacy. As seen in the accompanying table, it may also have a variety of implications and downsides.

Table 1: The Advantages and Disadvantages of Various Machine Learning Models

Model	Abilities	Limitations	Optimization Methods
Artificial Neural Network	It could be able to deal with nonlinear data and also has a strong fitting ability	Prone to overfitting, Prone to become stuck in a local optimum, Time-consuming model training	Adopted improved optimizers, activation functions, and loss functions
Support Vector Machine	It learns useful information from a small train set; Strong generation capability	It is difficult to perform well on big data or multiple classification tasks; Sensitive to kernel function parameters	Optimized parameters by particle swarm optimization (PSO)
K-Nearest Neighbor	Applicable to massive data; Suitable to nonlinear data, it also trains quickly and is robust to noise reduction in data.	Long test times, low accuracy for the minority class, and K parameter sensitivity	It used trigonometric inequality to shorten comparison times and PSO to optimize the settings.
Naïve Bayes	Resilient to noise; capable of progressive learning	Poor performance with attribute-related data.	Import latent variables to weaken the independent presumption.
Linear Regression	unassuming, quickly trainable; Features that scale automatically	perform poorly on nonlinear data; Apt to oversizing	Import regularization to prevent overfitting.

Decision Tree	It makes strong interpretations and automatically selects features.	Classification results tend to be in the majority; disregard data correlation	Balanced datasets with SMOTE; Introduced latent variables
K-means	Simple, can be trained rapidly; Strong scalability; Can fit to big data	On nonconvex data, it doesn't perform well; being initialization-sensitive; A K-parameter sensitivity	It requires an improved initialization method.

Source: [8, 2, 16, 10]

Artificial Neural Network (ANN)
The ANN was created to simulate human brain activity. This structure consists of an input layer, an output layer, and one or more hidden layers. There is perfect connectivity between units in an adjacent stratum. It can potentially approximate any function and has a very large number of units, which makes it exceptionally good at fitting functions, especially nonlinear ones. Due to the complex model structure, training ANNs requires time [7]. Incredibly, the models were trained using the back-propagation method, which can't be used to train deep networks. As a result, it is included in machine learning models, as shown in the image above.

Support Vector Machine (SVM)
The SVM technique is used to locate a max-margin separation hyperplane in the n-dimensional feature space. The separation hyperplane is determined by a minimal number of support vectors; hence it can be used to achieve satisfactory results even with small training sets. However, noise close to the hyperplane can affect SVMs. SVMs are good at resolving linear problems. For nonlinear data, kernel functions are typically used. A kernel function changes the original space into a new space to separate the original nonlinear data. The Kernel of SVM approaches are commonly used by SVMs as well as other machine learning techniques for prediction [16].

K-Nearest Neighbor (KNN)
It is based on many suppositions. If most of its neighbors also fall into that class, there is a substantial likelihood that the sample will as well. The only neighbors that matter for the classification result are the top k nearest neighbors. The parameter k has a significant impact on the performance of the KNN model. As k decreases, the likelihood of overfitting rises, complicating the model. On the other hand, as k rises, the model gets simpler and has a lesser capacity to fit the data [16].

Naïve Bayes
The conditional probability and the attribute independence hypothesis form the foundation of the Naive Bayes machine learning algorithm. The Nave Bayes classifier determines the conditional probability for several classes in each sample [18]. The sample is placed in the class with the highest probability. The conditional probability formula is computed as shown below.

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

If the attribute independence hypothesis is true, the Naive Bayes algorithm yields the best outcome. The Nave Bayes algorithm fares poorly on attribute-related data because, unfortunately, that hypothesis is difficult to prove in practice.

Logistic Regression (LR)
A particular kind of logarithm linear model is the LR [16]. It uses a parametric logistic distribution to calculate the probabilities of various classes, as illustrated below.

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

In where k = 1, 2,.., K 1. The class with the highest probability includes sample x. An LR model is simple to build, and model training is effective. However, LR's usefulness is constrained by its inability to handle nonlinear data well.

Decision Tree
The decision tree algorithm classifies data according to a set of criteria [3]. Since the model looks like a tree, it can be understood. It can automatically remove redundant and pointless characteristics from the sample dataset. The learning process includes feature selection, tree generation, and tree trimming. When training a decision tree model, the method selects the most appropriate features for each child node from the root node. The decision tree is one of the basic classifiers used in machine learning. Complex algorithms like random forest and extreme gradient boosting (XGBoost) employ several decision trees.

Clustering

This grouping is carried out in accordance with similarity theory, which places highly comparable data into the same clusters and less comparable data into other clusters [3]. Unlike classification, clustering is a form of unsupervised learning. Clustering algorithms' minimum data set needs are since they don't require prior knowledge or labeled data. However, while using clustering algorithms to find attacks, external data must be taken into consideration.

K-means is a popular clustering method, where K is the number of clusters and means is the average of the attributes. With the K-means method, distance serves as the benchmark similarity metric. With less space between them, there is a greater chance that two data objects will be grouped in the same cluster. K-means performs admirably on linear data, but on nonconvex data, it falls short of perfection. Further, the K-means algorithm is affected by the initialization condition as well as the value K. Numerous replicated tests must be performed to establish the correct parameter value [4].

Ensembles and Hybrids

Each machine learning algorithm has a different set of benefits and drawbacks. It makes sense to combine many ineffective classifiers to produce a strong classifier. Using ensemble methods, several classifiers are trained, and the classifiers then cast votes to decide the final results [5]. The various processes that make up hybrid techniques each need a classification model. Due to the fact that ensemble and hybrid classifiers frequently beat solo classifiers in classification problems, more academics are beginning to study them. The crucial decisions are which classifiers to combine and how to combine them.

Natural Gas Prediction

This fuel is not refined. In comparison to refined fuel oils like gasoline and diesel, natural gas is better adapted to reducing environmental pollutants [14]. As a fossil fuel, natural gas is one of them. It is environmentally friendly to use natural gas to supply energy requirements for industry, transportation, and other uses. Between a fifth and a third of the energy used worldwide is consumed by buildings. Currently, natural gas accounts for almost a third of the energy used in residential buildings in Europe [2]. Energy is necessary for the strategic development of civilizations and economies. Predicting energy consumption has become one of the primary concerns in many countries due to the exponential rise in global energy demand over the past few decades. About 34.7% of the world's energy is used for residential and commercial purposes [17]. Utilizing natural gas has been promoted as a way to improve energy security and lessen global environmental deterioration [5].

Data Mining Approach

It is a technology that combines methods from machine learning, statistics, and database systems to extract and spot patterns in vast volumes of data. Data mining methods are increasingly being used to anticipate the daily load on the natural gas system [26]. Creating models using preprocessed or existing data to discover patterns that the properties of the data set have revealed is the fundamental goal of data mining. By highlighting the connections between the attributes, some of the patterns act as explanations. Other modes anticipate future values for specific traits based on the most current information. Association patterns, prediction patterns, clustering patterns, and sequencing patterns are the four types of patterns that data mining is commonly used to identify [3].

Application of Machine Learning in Building Natural Gas Consumption Forecast

A computational system known as a machine learning model uses statistical and mathematical algorithms to discover patterns and correlations from data [15]. The model may predict or decide based on fresh, unobserved data after being trained on a dataset, which is a collection of examples or occurrences. Machine learning models come in a variety of forms, such as reinforcement learning models, unsupervised learning models, and supervised learning models. Supervised learning models are used for tasks like classification, regression, and prediction. They are trained on labeled data with the goal variable known. Unsupervised learning models are used for tasks like clustering and dimensionality reduction. They are trained on unlabeled data where the target variable is unknown. For tasks where an agent interacts with its environment and learns to adopt behaviors that optimize a reward signal, reinforcement learning models are utilized [13].

The study by Faruk *et al.* [1] employed machine learning algorithms to forecast the consumption of natural gas in the province of Istanbul. The results show that the support vector regression (SVR) technique is much superior to the artificial neural network (ANN) technique for time series forecasting of natural gas consumption, offering more accurate and dependable results with fewer prediction errors [16]. Due to the data format, consumer consumption patterns, and consumption patterns throughout many periods, this study may very well be used as a valuable benchmarking study for many developing countries.

The study by Šebal [7] offers an analysis of the approaches and tools used to forecast natural gas demand. The paper offers a comprehensive synthesis and analysis of the most recent research on natural gas consumption projections for residential and commercial use. The results show that while neural networks are the most popular, genetic algorithms, support vector machines, and Adaptive Network-based Fuzzy Inference System (ANFIS) are the most accurate approaches for estimating natural gas demand. The two most frequently used input variables are past natural gas consumption data and meteorological data, and predictions are most frequently generated on a daily and annual level on a national area level.

Natural gas is one of the main sources of energy on the entire planet, according to a study by Laib *et al*. [4]. By using a Multi-Layered Perceptron (MLP) neural network as a nonlinear predicting monitor, it presents a novel hybrid prediction model that addresses the weakness of the two-stage method [17]. The approach offers a fresh, useful functionality. In particular, during seasons with high consumer demand and consumption, its approximations of the next day's consumption profile significantly improved the results of prediction performance.

In their study, Bala and Shuaibu, [18] employed machine learning and hybrid techniques to forecast the energy consumption of the United Kingdom. While machine learning techniques like neural networks and support vector regression beat dynamic regression models, the seasonal hybrid model outperformed time series and machine learning models even when combination forecasts underperformed other models.

Energy consumption modeling for the residential and commercial sectors was also measured using machine learning. These models are developed based on some factors, including population, natural gas and electricity costs, gross domestic product, and the share of renewable energy sources in total energy consumption. The results of the three machine learning algorithms show that by 2040, Iranian residential and commercial energy consumption will be 76.97, 96.42, and 128.09 Mtoe, respectively. According to the findings, Iran must develop and implement new policies to increase the share of renewable energy sources in overall energy consumption [17].

The study by Qiao *et al.* [19] used a few hybrid algorithms that took their cues from nature to anticipate monthly natural gas use. The study found that invasive weed optimization (IWO), when compared to other hybridization procedures, can be a more precise prediction network for improving the MLP method's error performance. The anticipated results show that integrating optimization strategies may improve prediction accuracy and increase the multi-layer perceptron's (MLP) output.

Artificial neural networks (ANN) were also used to forecast ultimate natural gas usage. Speculate about how natural gas will be used in the future. For Myanmar, actual recorded consumption data from 1990 to 2015 are used. The most recent five years' worth of data (2011-2015) are combined with the prior years' data for testing. The years (1990–2010) are used for instruction. Data on the population, GDP, and other factors that influence the final use of natural gas are used to develop the model. The training model is reliable (MSE) with a minimum error rate of 0.005 Mean Squared Error. Myanmar's yearly natural gas consumption is projected using the developed ANN model [20].

To examine the effects of various distribution network operator (DNO) tactics and grid-specific factors, grid charges in 57 German urban distribution grids were modeled. The effects on DNO business models are determined by assessing the total operating expenses and associated grid charges for various scenarios and tactics. Grid prices may rise as a result of the substitution of gas-bound equipment, which could create a self-reinforcing feedback loop and result in grid defection, depending on the policies of the DNO and the pattern of consumers leaving the grid.

Monthly power usage ratings are classified precisely for a region with more than 16,000 residential buildings based on publicly available information. The first step is to employ data mining techniques to find and collect any hidden patterns of electricity usage in the data. After the clustering analysis has been sent to the particle swarm optimization-K-means method, the cluster centers are utilized to divide the amount of electricity usage. The last idea is a practical, effective classification technique that uses a support vector machine model to optimize performance [21]. The findings show that the new model's accuracy and F-measure, which both significantly beat traditional approaches, reach 96.8% and 97.4%, respectively.

The purpose of this work is to solve the issues raised by Shapi *et al.* [16] on the development of an energy usage forecast model using Microsoft Azure's cloud-based machine learning platform. Two tenants of a commercial building are utilized as a case study to focus on reallife application in Malaysia. Before being used to train and test models, the acquired data is inspected and prepared. RMSE, NRMSE, and MAPE indicators are used to compare the effectiveness of each technique. The results of the experiment demonstrate that the distribution of energy use varies depending on the renter.

In order to forecast the consumption of natural gas in Bahçeşehir, Istanbul, several effective machine learning approaches were employed, and their effectiveness was assessed. By a mean absolute error of 0.04 for each, XGBoost ultimately outscored MLP, Forest Regression, and Linear Regression. XGBoost is a superior model for prediction because it is extremely scalable, effectively reduces computation time, and optimally utilizes memory [22]. Accurate projections stop economic loss and maintain supply and demand equilibrium.

Computed the amount of natural gas used in data centers [6]. The data center energy scheduling model is constructed by taking into account the service level of the data center, with the assumption that electricity and gas are used as energy providers of energy supply and energy consumption from two viewpoints. The data center's schedule calculation model is the lowest. The data center energy supply scheduling model is the most effective one. The timeline is imitated using the particle swarm methodology. The findings demonstrate that adding natural gas as an alternative energy source can drastically lower the data center's overall energy usage while taking into account the degree of service delay.

The work of Chen and Zhang [31] presents a reliable method for estimating intraday load for the natural gas flow state space model. The results of the trials demonstrate that the model's relative error and load forecasting accuracy both reached 0.026 and 98.5%, respectively, resolving the issue of processing the long-term gathered historical data of gas intra-day load. Along with encouraging the qualitative study and the measurement of factors impacting intraday load, the calculation's data input was also decreased by 33.6%.

Machine learning was picked after looking through the available models. In particular, genetic algorithms (GA) and artificial neural networks (ANN) were combined. Their suggested models were put to use in a real-world testbed. They took advantage of CompactRIO to put ANNs into practice. The proposed models are trained and validated using actual data from a PV system and SB electrical appliances [23]. The model is therefore strongly suggested as a guide for researchers willing to implement real-world SB testbeds and investigate machine learning as a promising venue for energy consumption prediction and scheduling, despite the model's modest prediction accuracy, which is caused by the small size of the data set [24].

Check out data-driven prediction models for estimating natural gas prices that are based on well-known machine learning techniques, such as Gaussian process regression, support vector machines (SVM), and artificial neural networks (ANN) (GPR) [5]. We used monthly Henry Hub natural gas spot price data from January 2001 to October 2018 to assess the model and train it using the cross-validation technique. The findings indicate that these four machine-learning techniques have varying degrees of success in predicting natural gas prices. Overall, ANN generates predictions that are more accurate than those made by SVM, GBM, and GPR. This has shown that machine learning classifiers can predict outcomes for many tasks with accuracy.

The research could be used to considerably enhance natural gas consumption forecast systems, despite its limitations due to the small number of articles it looked at. Accurate estimates of natural gas consumption are essential for supply-demand equilibrium and investment goals. Accurate and exact forecasts prevent economic loss and keep supply and demand in balance.

Related Works

Table 2: Summary of various methods used in the literature

Reference	Proposed Methods	Limitation/drawback
[1]	An artificial neural network method (ANN), support vector regression (SVR), and multiple linear regression (MLR)	The proposed model does well but such models tend to work well on a specific dataset but do not generalize well
[10]	VM (Support Vector Machine), LR (Linear Regression), ELM (Extreme Learning Machine), and ANN (Artificial Neural Network)	Model accuracy could be biased. The performance of the technique could be lowered if the dataset contains too much noise.
[16]	Artificial neural networks (ANN), k-nearest neighbor (KNN), and support vector machines (SVM)	This technique allows bogus accounts, which often have zero or one post, to occasionally get under Instagram's restrictions.
[17]	Artificial neural networks that use nonlinear autoregressive with exogenous input, multiple linear regression, and logarithmic multiple regression	Using Machine Learning to Model Energy Consumption in the Residential and Commercial Sectors
[18]	Techniques for modeling data using machine learning, dynamic regression, time series, and combination models	The proposed project focuses on categorizing articles as legitimate or fraudulent without taking into account their sources.
[19]	The genetic algorithm (GA-MLP), the dragonfly algorithm (DA-MLP), the evolution strategy (ES-MLP), the invasive weed optimization (IWO-MLP), and the imperialist competitive algorithm (ICA-MLP).	The proposed model was unable to estimate the monthly natural gas consumption (NGC) utilizing one overarching strategy.
[2]	investigating the impact of grid-specific variables and various distribution network operator (DNO) methods on grid charges	Problems with interpretability and computational effectiveness
[21]	support vector machine (SVM)	Does not take into account the diversity of agents
[22]	The following regression methods are used: gradient boosting (XGBoost), multilayer perceptron with back propagation (MLP), linear regression (LR), and random forest regression (RFR).	It is intended to deal with the issue of predicting different product demands for key distribution warehouses, although it has some requirements for system complexity
[23]	Artificial Neural Networks (ANN)	High computational cost when training large datasets
[23]	traditional classifiers for machine learning	It was mainly concerned with attaining high levels of precision and accuracy without any computing limitations.
[24]	The use of Genetic Algorithms and Artificial Neural Networks (ANN)	Selecting reasonable parameters is very stimulating, which may be why it isn't very satisfactory

[25]	support vector machine (SVM) and the improved artificial fish swarm algorithm (IAFSA)	Large datasets are a major cause of slow training. The performance of the model is decreased when the dataset has a lot of noise; for example, target classes may easily overlap. A small change to the dataset could have a significant impact on the forecast.
[3]	Data Mining methodology	Effectively encouraging qualitative analysis and the identification of intraday load influencing factors reduced the amount of data utilized in the calculation as well.
[4]	Long Short-Term Memory (LSTM) Recurrent Neural Networks	The proposed model was unable to account for the variation in gas demand and consumption, particularly around holidays, using a single, overarching strategy.
[5]	Gradient Boosting Machines (GBM), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Gaussian Process Regression (GPR)	Dataset size is a major cause of slow training. Target classes may easily overlap when there is a lot of noise in the dataset, which lowers model performance. The forecast could be significantly affected by a small change in the dataset
[6]	Scheduling of Electric Energy and Natural Gas Energy Consumption in Data Centers under Optimal Condition.	Even though it is superior, it is distinguished by high energy consumption, which is the main cause of the data center's expensive running expenses.
[7]	Analysis of Prediction Methods and Techniques for Natural Gas Consumption	Despite the study's limitations due to the small number of articles it evaluated, it might still be used to significantly enhance natural gas consumption forecast systems.
[8]	Extreme Learning Machine (ELM), Adaptive Network-based Fuzzy Inference System (ANFIS), and Conventional Machine Learning	perform poorly while performing numerous categorization tasks or using huge data; Adaptable to the kernel function parameters
[9]	A review of energy consumption prediction techniques using machine learning	Classification results tend to be in the majority; disregard data correlation

DISCUSSION, CHALLENGES AND RESEARCH OPPORTUNITIES

Residential natural gas demand and consumption forecasting has demonstrated to have a lot of potential. To further increase the accuracy and efficacy of these techniques, however, many issues and research possibilities must be addressed [27]. The accessibility and caliber of the data represent one of the major obstacles. Machine learning algorithms need historical weather and natural gas usage data to be trained, but this data might be sparse or unreliable, especially at the household level [20]. Future research could look into using social media data or data from smart meters to supplement traditional data sources to get around this problem.

The diversity of households and their patterns of energy usage present another difficulty

[15, 22]. It is possible that machine learning models that are effective for one task may not always be able to generalize to another with varied patterns of energy consumption. Future studies could investigate the use of personalized or adaptive machine-learning models that can take into account unique family traits and behavior to address this difficulty [28, 29]. Additionally, a significant difficulty is the interpretability and transparency of machine learning models for predicting natural gas demand because end users may not always be aware of or comprehend the aspects that influence the projections. Future research might examine the application of explainable AI techniques to improve the readability and interpretability of machine learning models, hence enhancing their reliability and usefulness.

The incorporation of machine learning models into energy management systems to give real-time feedback and recommendations to end users represents another significant area for future research [14]. Households may be able to optimize their energy consumption practices and lower their costs and natural gas usage as a result. In summary, there are many obstacles and areas for future research for machine learning algorithms to forecasting residential natural gas demand and consumption [21, 30]. Taking on these issues and following these research opportunities may result in more precise and efficient machine-learning models that assist consumers and energy suppliers in maximizing their use of natural gas while lowering prices and environmental impact.

CONCLUSION AND FUTURE DIRECTION

In conclusion, machine learning techniques could considerably increase the efficacy and accuracy of predicting natural gas demand and consumption in residential settings. Energy suppliers may improve their pricing and supply chain strategies with the support of precise natural gas demand and consumption predictions, and can also give homeowners more individualized and accurate energy usage advice. Despite the difficulties that come with using machine learning to predict natural gas demand and consumption, including data accessibility and quality, interpretability, and the heterogeneity of households, several research opportunities can aid in enhancing the precision and efficacy of these methods. These possibilities include the application of personalized or adaptive machine learning models, the integration of smart meter data, and the incorporation of machine learning models into energy management systems. Machine learning algorithms for predicting natural gas demand and consumption are expected to take a promising trend in the future. We may anticipate seeing more precise and potent models that can assist homes and energy suppliers in better managing their natural gas usage and prices as machine learning techniques continue to progress and become more widely available. Finally, this can assist a future with more sustainable energy sources and lessen the negative effects of natural gas consumption on the environment.

REFERENCES

1. FARUK, O.B., CAYIR, E., TATOGLU, GOKCIN, P. & ZAIM, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul, *Energy Economics*, **80**: 937-949, Doi: 10.1016/j.eneco.2019.03.006.
2. THEN, D., SPALTHO, C., BAUER, J., KNEISKE, T.M. & BRAUN, M. (2020). Impact of Natural Gas Distribution Network Structure and Operator Strategies on Grid Economy in Face of Decreasing Demand, 2020.
3. CHEN, L. & ZHANG, J. (2022) A Forecast Model of City Natural Gas Daily Load Based on Data Mining, *Science Program*, 2022, Doi: 10.1155/2022/1562544.
4. LAIB, O., TAREK, M. & MIHAYLOVA, L. (2019). Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks, *Energy*, **177**: 530-542, 2019, doi: 10.1016/j.energy.2019.04.075.
5. GAS, N., DRIVEN, D. PRICE, S., USING, M. & LEARNING, M. (2019). Data Driven Natural Gas Spot Price Prediction Models Using Machine Learning Methods, 2019.
6. LIU, X., HOU, G. & YANG, L. (2022). Dynamic Combined Optimal Scheduling of Electric Energy and Natural Gas Energy Consumption in Data Center, *Discrete Dynamics in Nature and Society.*, vol. 2022, doi: 10.1155/2022/3917170.
7. ŠEBAL, J.D. (2019). Analysis of Methods and Techniques for Prediction of Natural Gas Consumption: A Literature Review Analysis of Methods and Techniques for Prediction of Natural Gas Consumption: *A Literature Review*, no. June 2019, doi: 10.31341/jios.43.1.6.
8. BARIK, S.S. & NAYAK, S. (2021). Electricity Consumption & Prediction Using

Machine Learning Models. India, Odisha, and management Centurion University of Technology and Management, Department of Computer Science and Engineering, Sasmita Kumari Nayak Assistant Professor, September 2021.

9. MOSAVI, A. & BAHMANI, A. (2019). Energy Consumption Prediction Using Machine Learning; A Review Energy consumption prediction using machine learning ; a review, no. March, 2019.
10. LI, J., GUO, Y., ZHANG, X. & FU, Z. (2021). Using Hybrid Machine Learning Methods to Predict and Improve the Energy Consumption Efficiency in Oil and Gas Fields, *Mobile Information Systems*, vol. 2021, doi: 10.1155/2021/5729630.
11. ABBA, S., AJUJI, M. & MUHAMMAD, A. N. (2022). Performance Evaluation and Analysis of Supervised Machine Learning Algorithms for Bitcoin Cryptocurrency Price Forecast, no. 16, pp. 28–42.
12. CHOUBEY, S. & KARMAKAR, G.P. (2021). Artificial intelligence techniques and their application in the oil and gas industry, *Artificial Intelligence Review*, **54**(5): 3665–3683, doi: 10.1007/s10462-020-09935-1.
13. LI, H., YU, H., CAO, N., TIAN, H. & CHENG, S. (2021). Applications of Artificial Intelligence in Oil and Gas Development, *Archives of Comput. Methods in Engineering*, **28**(3): 937–949, doi: 10.1007/s11831-020-09402-8.
14. A. SIRCAR, K. YADAV, K. RAYAVARAPU, N. BIST, & H. OZA, (2021). Application of machine learning and artificial intelligence in oil and gas industry, *Petroleum Research*, **6**(4): 379–391, doi: 10.1016/j.ptlrs.2021.05.009.
15. LALMUANAWMA, S., HUSSAIN, J. & CHHAKCHHUAK, L. (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review, *Chaos, Solitons and Fractals*, **139**: 110059, 2020, doi: 10.1016/j.chaos.2020.110059.
16. SHAPI, M.K.M., RAMLI, N.A. & AWALIN, L.J. (2020) Energy Consumption Prediction by using Machine Learning for Smart Building: Case study in Malaysia. Development in the Built Environment., 2020, p. 100037, doi: 10.1016/j.dibe.2020.100037.
17. NABAVI, S.A., ASLANI, A., ZAIDAN, M.A., ZANDI, M., MOHAMMADI, S. &MOTLAGH, N.H. (2019). Machine Learning Modeling for Energy Consumption of Residential and Commercial Sectors, pp. 1-22, doi:10.3390/en13195171.
18. BALA, D.A. & SHUAIBU, M. (2022). Forecasting United Kingdom's energy consumption using machine learning and hybrid approaches, 2022, doi: 10.1177/0958305X221140569
19. QIAO, W., MOAYEDI, H. & ORIGINAL, W. (2020). Energy & Buildings Nature-inspired hybrid techniques of IWO, DA, ES, GA, and ICA, validated through a k-fold validation process predicting monthly natural gas consumption, vol. 217, 2020, doi: 10.1016/j.enbuild.2020.110023
20. YIN, K.S. & HTAY, S.S. (2011). Prediction of Natural Gas Final Consumption using Artificial Neural Networks, pp. 224–229, 2011.
21. CAI, H., SHEN, S., LIN, Q., LI, X. & XIAO, H.U.I. (2019). Predicting the Energy Consumption of Residential Buildings for Regional Electricity Supply-Side and Demand-Side Management, *IEEE Access*, **7**(3): 30386–30397, doi: 10.1109/ACCESS.2019.2901257.
22. AHMED, S. MADANIAN, S. & MIRZA, F. (2021). 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, no. February 2021.
23. COMPUT, J.P.D., GARCA-MARTN, E., FAVIOLA, C., RILEY, G. & GRAHN, H. (2019). Estimation of energy consumption in machine learning. *Journal of Parallel Distributed Computing,* **134:** 75–88, doi: 10.1016/j.jpdc.2019.07.007.
24. BOURHNANE, S., RIDUAN, M., RACHID, A., KHALID, L., DINE, Z. & ELKAMOUN, N. (2024) Machine learning for energy consumption prediction and scheduling in smart buildings, no. 2020, 2024.
25. GAO, F. & X. SHAO, (2021). Forecasting annual natural gas consumption via the application of a novel hybrid model, pp.

21411–21424, 2021.

26. PELLETIER, J.P., JAMSHIDI, A. & MARTEL-PELLETIER, J. (2019) Machine-learning-based patient-specific prediction models for knee osteoarthritis, *Nature Reviews Rheumatology*, 15**(**1): 49–60, doi: 10.1038/s41584-018-0130-5.
27. LAWAL, M., MUSA, K., YAU, A. & UMAR, F. (2021) Science Direct 10th International Young Scientists Conference on Computational Science A Review on Deep Learning with Focus on Deep Recurrent Neural Network for Electricity Forecasting in Residential Building, *Procedia Computer Science*, **193(**1): 141–154, doi: 10.1016/j.procs.2021.10.014.
28. SCHWARTZ, A.J., CLARKE, H.D., SPANGEHL, M.J., BINGHAM, J.S., ETZIONI, D.A. & NEVILLE, M.R. (2020). Can a Convolutional Neural Network Classify Knee Osteoarthritis on Plain Radiographs as Accurately as Fellowship-Trained Knee Arthroplasty Surgeons?, *Journal of Arthroplasty*, **35**(9): 2423–2428, doi: 10.1016/j.arth.2020.04.059.
29. BIN KWON, S., KU, Y., UK SOO HAN, H., LEE, M.C., KIM, H.C. & RO, D.H. (2020). A machine learning-based diagnostic model associated with knee osteoarthritis severity, *Science Reports*, **10**(1): 1-8, doi: 10.1038/s41598-020-72941-4.
30. LEUNG, K., ET AL., (2020). Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: Data from the osteoarthritis initiative, *Radiology*, **296(**3): 584-593, doi: 10.1148/radiol.2020192091.
31. CHEN, L. & ZHANG, J. (2022). A Forecast Model of City Natural Gas Daily Load Based on Data Mining, *Hindawi Scientific Programming*, Volume 2022, Article ID 1562544, 14 pages, https://doi.org/10.1155/2022/1562544